Maria-Gabriella Di Benedetto
Andrea F. Cattoni
Jocelyn Fiorina
Faouzi Bader
Luca De Nardis   *Editors*

Cognitive Radio and Networking for Heterogeneous Wireless Networks

Recent Advances and Visions for the Future

Springer

Signals and Communication Technology

Maria-Gabriella Di Benedetto • Andrea F. Cattoni •
Jocelyn Fiorina • Faouzi Bader • Luca De Nardis
Editors

Cognitive Radio and Networking for Heterogeneous Wireless Networks

Recent Advances and Visions for the Future

Springer

Editors

Maria-Gabriella Di Benedetto
Electronics and Telecommunications
Sapienza University of Rome
Rome, Italy

Andrea F. Cattoni
Electronic Systems
Aalborg University
Aalborg, Denmark

Jocelyn Fiorina
Telecommunications
Supeléc
Paris, France

Faouzi Bader
École Supérieure d'Électricité
Rennes, France

Luca De Nardis
Sapienza University of Rome
Rome, Italy

Foreword

COST Action IC0902 "Cognitive Radio and Networking for Cooperative Coexistence of Heterogeneous Wireless Networks" proposed coordinated research in the field of cognitive radio and networks; research activities were carried out within COST Action IC0902 between December 2009 and December 2013. Research directions were set at the beginning of the Action, and proved to have a scope and vision as wide as a 4-year research program would require.

The main objective of the Action was to integrate the cognitive concept across all layers of communication systems, with the aim of defining a European platform for cognitive radio and networks. The cognitive concept was applied to coexistence between heterogeneous wireless networks sharing the electromagnetic spectrum for maximum efficiency in resource management. The Action promoted coordinated research in the field of cognitive radio and networks in order to synergize the several trends that were in place in European research centers and consortia, consisting in introducing cognitive mechanisms at different layers of the communications protocol stack.

This Action went beyond the above trends by integrating the cognitive concept across all layers of system architecture, in view of joint optimization of link adaptation based on spectrum sensing, resource allocation, and selection between multiple networks, including underlay technologies. The cross-layer approach provided a new perspective in the design of cognitive systems, based on a global optimization process that integrated existing cognitive radio projects, thanks to the merge of a wide-range of expertise, from hardware to applications, provided by over 30 academic and industrial partner institutions. The final result was the definition of a European platform for cognitive radio and networks; algorithms and protocols for all layers of the communications stack were designed, as well as a set of standard interfaces.

Beyond scientific achievements, COST Action IC0902 had a strong impact on the scientific community under two aspects: formation of young researchers and interaction with industrial consortia. In terms of spin-off of new projects, about 20 new EU and national RTD programs were generated. An inter-disciplinary approach

was pursued and networking with experts coming from research areas such as probabilistic models, mathematics and economics was incentivized.

Action IC0902 involved 17 COST countries and 7 institutions from 6 non-COST countries, and more than 200 researchers, with over 50 % of Early Stage Researchers. Early Stage Researchers played a central role in the Action; four training schools attended by more than 350 young researchers, as well as four yearly workshops, where Early Stage Researchers presented over 160 technical contributions, and two mini-workshops focusing on platforms and learning techniques, were organized.

This book is targeted to academic researchers, designers and graduate students wishing to enhance their understanding of cognitive radio and networks; the book includes and illustrates the most salient research results of COST Action IC0902's 4-year collective efforts, and forms a scientific selected outcome of the Action.

It is a great pleasure to introduce this book and to have hereby the chance to thank and congratulate all IC0902 participants, especially the authors of the different chapters of the book, as well as, in particular, Andrea Fabio Cattoni for making this book possible.

Rome, Italy   Maria-Gabriella Di Benedetto
February 2014   Chair of the COST Action IC0902

Preface

The aim of this book is to present, in a tutorial perspective, the visions and trends for future wireless networks and technologies based, inspired, and learned from Cognitive Radio (CR) investigations that have been carried out within the framework of the 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 programs supporting cooperation, sharing, and dissemination among scientists and researchers across Europe.

This book has been written by leading researchers, that are first-hand experts in a wide range of fields and disciplines, from both Academia and industrial players. The main purpose of the manuscript is to provide scientists, researchers, wireless communication network engineers and professionals, and graduate students with a widely comprehensive tutorial tool for understanding and developing cognitive radio network (CRN) technology components, based on the latest state of the art in the field.

The book will allow the reader to access both fundamental and advanced knowledge spanning from the design of CRN air interfaces to network architectures and algorithms, following a bottom-up approach. Furthermore, the book includes practical help in setting up essential investigation tools such as simulators and experimental test beds, in the form of direct step-by-step how-to.

Topics covered by the book include filter bank multicarrier systems that try to incorporate coexistence properties starting from the design of the CR air interface. Furthermore, recent advances in signal processing are presented in the context of interference mitigation. The reader will also get access to insights on how to design the link layer of CR, including specific highlights on essential features such as spectrum sensing and dynamic spectrum access. Climbing up in the traditional layered architecture, the book will also provide trending visions on how to evolve the internet architecture and protocols in dynamically adaptive, cognitive-oriented ones. An interesting chapter will provide also regulatory perspectives on the possible usage of the spectrum by CRN players, including economical and market reflections. Finally the book will dive into the practical aspects that a researcher

or an engineer can face on the everyday job: how to simulate or develop CRN systems. Three chapters cover the most hot topics in the field, using open-source widely available tools, such as software simulations, software defined radio tools and developments, and embedded systems.

A brief description of each chapter is as follows.

In Chap. 1 by Medjahdi et al. the authors provide a tutorial overview on multicarrier techniques, which are offering flexible access to possible new spectrum opportunities. Indeed, OFDM, which is the most commonly used multicarrier technique, has been adopted in IEEE 802.22 standard for unlicensed wireless regional area network (WRAN) considering cognitive communications on the unused TV bands. Unfortunately, OFDM presents some weaknesses. In fact, the redundancy, caused by the insertion of the cyclic prefix mandatory part of the transmitted OFDM symbol, reduces the useful data rate. Furthermore, the poor spectral localization of the OFDM subcarriers due to the large sidelobes induces not only an additional spectral loss but also interference problems with unsynchronized signals. These shortcomings have stimulated the research of an alternative scheme that can overcome these problems. An enhanced physical layer based on the filter bank processing called filter bank-based multicarrier (FBMC) technique has been proposed in various works. A major improvement in spectrum efficiency and better flexibility for system coexistence can be achieved by FBMC thanks to the use of spectrally well-localized waveforms. The advantages and the drawbacks of the classical OFDM are first discussed in the chapter. Next, the authors will introduce the different schemes of FBMC systems: Filtered multitone (FMT), cosine modulated multitone (CMT) and staggered multitone (SMT, also called FBMC-OQAM). After introducing a theoretical background of the FBMC transmission, the polyphase implementation of the filter bank transceiver and the prototype filter design are reviewed. Finally, the chapter highlights the interest of FBMC spectrum sharing and investigate coexistence issues.

Interference alignment (IA) has been widely recognized as a promising interference mitigation technique since it can achieve the optimal degrees of freedom in certain interference limited channels. This topic is then addressed by Sharma et al. in Chap. 2. IA allows in fact the coexistence of two heterogeneous wireless systems in an underlay cognitive mode. The main concept behind this technique is the alignment of the interference on a signal subspace in such a way that it can be filtered out at the non-intended receiver by sacrificing some signal dimensions. This chapter starts with an overview of the IA principles, the degree of freedom (DoF) concept, and the classification of existing IA techniques. Furthermore, this chapter includes a discussion about IA applications in CR networks. Moreover, a generic system model is presented for allowing the coexistence of two heterogeneous networks using IA approach while relevant precoding and filtering processes are described. In addition, two important practical applications of the IA technique are presented along with the numerical results for underlay spectral coexistence of (i) femtocell-macrocell systems, and (ii) monobeam-multibeam satellite systems. More specifically, an uplink IA scheme is investigated in order to mitigate the interference of femtocell user terminals (UTs) towards the macrocell base station (BS) in the

spatial domain and the interference of multibeam satellite terminals towards the monobeam satellite in the frequency domain.

Chapter 3 by Yilmaz et al. will provide a detailed technical insight into latest key aspects of cooperative spectrum sensing. The authors focus on fusion strategies, quantization enhancements, effect of imperfect reporting channel, cooperative spectrum sensing scheduling, and utilizing cooperatively sensed data via radio environment map (REM). The sharing of local observations between the secondary users and the fusion center is one of the most crucial factors that determines the performance of cooperative sensing. The detection performance is determined by the quality of local observations and the quality of the information received by the fusion center. Therefore, the number of quantization bins, the number of bits sent for sensing reports, the global decision logic, the imperfections in the reporting channel, and the erroneous reports due to malfunctioning or malicious secondary devices affect the system performance. Furthermore, there are many channels to sense while the cooperating nodes are few, therefore coordinating the sensing nodes for detecting high quality channels is necessary. Cooperative sensing scheduling concentrates on the scheduling of cooperative nodes and the channels to be sensed. This chapter also focuses on the energy consumption problem that becomes more severe if the users are mobile. The components of energy consumption dedicated to cooperative sensing are analyzed and optimal and sub-optimal (but efficient) sensing scheduling mechanisms are discussed in order to reduce the sensing energy consumption of the network. Once the spectrum has been sensed cooperatively, the outcomes can be utilized via REM, which can be considered as a crucial part of the cognitive engine located at the network. The sensed information may also play a crucial role in the generation of REM. Hence, this chapter also focuses on how the sensing measurements could be utilized for REM construction.

The endeavor of categorizing the existing cognitive-MAC (C-MAC) protocols requires definition of general classification frameworks or layouts that merge most of the aspects of the protocols in a single unified presentation. Chapter 4 by Gavrilovska et al. introduces the C-MAC cycle as a general classification and systematization layout for C-MAC protocols. The C-MAC cycle originates from the idea that the MAC layer in spectrally heterogeneous environments should provide support for three generic technical features: radio environmental data acquisition; spectrum sharing; and control channel management. The inclusion of these generic technical features is necessary in CRNs for improving the network performance and achieving spectrum efficiency gain while providing maximal level of protection for the primary system. This chapter will present extensive survey on the state-of-the-art advances in C-MAC protocol engineering by reviewing existing technical solutions and proposals, identifying their basic characteristics and placing them into the C-MAC cycle, with emphasis on the modularity of the C-MAC cycle. It provides an overview of a large number of technical details concerning the three generic functionalities (i.e. the radio environmental data acquisition, the spectrum sharing and the control channel management) as the main building blocks of the C-MAC cycle. Three use cases (each in different generic functional group) illustrate the capabilities of the proposed C-MAC cycle layout. In more detail, the first use

case theoretically presents and practically evaluates cooperative spectrum sensing based on estimated noise power. The results illustrate the effect of estimating the noise variance on the detection capabilities of the majority voting and equal gain combining cooperative spectrum sensing strategies. The second use case presents advanced and computationally efficient horizontal spectrum sharing strategy for secondary systems based on node clustering and beamforming. Finally, the last use case presents and assesses a multiuser quorum-based multiple rendezvous strategy for control channel establishment in distributed CRNs.

In Chap. 5 by da Costa et al. the authors dive in the ever-growing demand for mobile broadband that is pushing towards the utilization of small cells, including metrocells, picocells and femtocells. In particular, the deployment of femtocells introduces significant challenges. Firstly, the massive number of expected femtocells cannot be deployed using the traditional planning and optimization techniques. This leads to uncoordinated deployment by the end-user. Secondly, the high density of femtocells, including vertical reuse, leads to very different inter-cell interference patterns than the ones traditionally considered in cellular networks. And last, but not least, the possibility of having closed-subscriber-groups aggravates the inter-cell interference problems. In order to tackle these issues the authors consider the implementation of some aspects of CR technology into femtocells, leading to the concept of cognitive femtocells (CFs). This chapter focuses on state-of-the-art techniques to manage the radio resources in order to cope with inter-cell interference in CFs. Different techniques are presented as examples of gradually increasing sophistication of the CFs, allowing for dynamic channel allocation, dynamic reuse and negotiated reuse based on information exchanged with neighbor cells.

Granelli et al. are instead moving the discussion from the wireless domain into the networking one in Chap. 6. As a matter of fact, the requirement to support an always increasing number of networking technologies and services to cope with context uncertainties in heterogeneous network scenarios leads to an increase of operational and management complexity of the Internet. Autonomous communication protocol tuning is then crucial in defining and managing the performance of the Internet. This chapter presents an evolutionary roadmap of communication protocols towards cognitive Internet in which the introduction of self-aware adaptive techniques combined with reasoning and learning mechanisms aims to tackle inefficiency and guarantee satisfactory performance even in complex and dynamic scenarios. In this chapter, the authors provide a survey and comparison between existing adaptive protocol stack solutions, reviewing the principles of cross-layer design as well as the agent-based and AI based self-configuration solutions. The fundamental principles of cognitive protocols, such as adaptation, learning, and goal optimization, are presented along with implementation examples.

Introducing cognitive mechanisms at the application layer is instead investigated by Boldrini et al. in Chap. 7 and it may lead to the possibility of an automatic selection of the wireless network that can guarantee best perceived experience by the final user. This chapter investigates this approach based on the concept of quality of experience (QoE), by introducing the use of application layer parameters, namely key performance indicators (KPIs). KPIs are defined for different traffic types based

on experimental data. A model for an application layer cognitive engine is presented, whose goal is to identify and select, based on KPIs, the best wireless network among available ones. An experimentation for the VoIP case, that foresees the use of the one-way end-to-end delay (OED) and the mean opinion score (MOS) as KPIs is presented. This first implementation of the cognitive engine selects the network that, in that specific instant, offers the best QoE based on real captured data.

Another important piece of information that CRNs can use for optimizing services is the location of the CR terminal. This aspect will be clarified by Chochliouros et al. in Chap. 8. Starting from a general survey of several among the critical capabilities and/or features characterizing CRNs, in the context of actual European standardization efforts, the chapter will present an overall and harmonized technical concept for future CR systems, especially by discussing several options affecting the future evolution of radio technologies and network architectures towards more flexible and reconfigurable CR systems, as the latter are expected to increase the efficiency of the overall spectrum usage by offering new sharing opportunities and thus to provide more flexibility to applications-services.

In Chap. 9 Georgakopoulos et al. will provide a trending vision on the architectural possibilities that are offered to wireless mobile broadband networks for jointly satisfying complex context of operations as well as system requirements like QoE (quality of experience) and energy and cost efficiency. Introduction of intelligence in the Cloud-RANs will lead towards this direction by providing the necessary decisions based on the received inputs. Cloud-RANs have the capabilities to adapt their network topology and resource allocation so as to realize environmental-friendly and cost-efficient solutions by moving elements of the legacy RAN to cloud-based infrastructures. The authors will try to provide, in the chapter, an indication on the elements of the approach as well as the identification of the benefits from such a concept.

The purpose of Chap. 10 by Doyle and Forde is instead to give a regulatory perspective on CR, that can be seen as a natural part of the roadmap for advanced communication systems and from this standpoint can be dealt with within the context of normal regulations. However one of the key and unique advantages of CR is that it is an enabler of spectrum sharing in its many forms. Hence the main part of this chapter is devoted to different regulations in spectrum sharing and the implications for CR. It looks at regulations which are in existence as well as emerging regulations. The chapter also provides a generic framework in which to place different sharing regimes.

Chapter 11 by Caso et al. is leading the discussion in the practical everyday issues of researchers and engineers. In fact, a widespread methodology for performance analysis and evaluation in communication systems engineering is network simulation. It is widely used for the development of new architectures and protocols. Network simulators allow to model a system by specifying both the behavior of the network nodes and the communication channels, and CR-related research activities have been often validated and evaluated through simulation too. Following this approach, this chapter presents an ongoing effort towards the development of a CR simulation framework, to be used in the design and the evaluation of protocols

and algorithms. OMNeT++, in combination with MiXiM framework functionalities, was chosen as the developing platform, thanks to its open source nature, the existing documentation on its architecture and features, and the user-friendly integrated development environment (IDE).

More and more researchers are entering the idea of research-oriented test beds. Unfortunately, it is very difficult for a wide number of research groups to start with their own set up, since the potential costs and efforts could not pay back in term of expected research results. Chapter 12 by Cattoni et al. provides a tutorial, first-hand overview on software defined radio solutions, which offer an easy way to communication researchers for the development of customized research test beds. While several hardware products are commercially available, the software is most of the times open source and ready to use for third party users. Even though the software solution developers claim complete easiness in the development of custom applications, in reality there are a number of practical hardware and software issues that research groups need to face, before they are up and running in generating results. With this chapter the authors will provide a tutorial guide, based on direct experience, on how to enter in the world of test bed-based research, providing both insight on the issues encountered in everyday development, and practical solutions. Finally, an overview on common research-oriented software products for SDR development, namely GNU radio, Iris, and ASGARD, will be provided, including how to practically start the software development of simple applications.

In Chap. 13 Šolc et al. describe their experiences with the design, deployment and experimental use of the LOG-a-TEC embedded, outdoor CR test bed, based on the VESNA sensor node platform. The authors will describe the choice of experimental low-cost reconfigurable radio frontends for LOG-a-TEC and discuss the potential capabilities of custom designs. The core part of this chapter will provide practical experiences with designing the embedded testbed infrastructure, covering topology design and performance evaluation of the management network as well as considerations in the choice of network protocols employed in the LOG-a-TEC testbed. Use cases using LOG-a-TEC test bed for performing experiments are also covered, one relevant to the investigation of coexistence of primary and secondary users in TV white spaces and the other addressing power allocation and interference control in the case of shared spectrum.

Finally, as Editors of this volume, we would like to express our gratitude to all the contributors for their help, support, and effort in making this book possible. Furthermore, we would like to thank the Springer editorial team for their continuous support along the publication process.

Roma, Italy   Maria-Gabriella Di Benedetto
Aalborg, Denmark   Andrea Fabio Cattoni
Paris, France   Jocelyn Fiorina
Rennes, France   Faouzi Bader
Roma, Italy   Luca De Nardis
February 2014

Acknowledgements

Contents

List of Contributors

George S. Agapiou Hellenic Telecommunications Organization S.A. (OTE), Maroussi, Greece

Stelios Agapiou Department of Informatics and Telecommunications, National Kapodistrian University of Athens, Athens, Greece

Marko Angjelicinoski Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje, Macedonia

Sergio Benco CSP – ICT Innovation, Turin, Italy

Stefano Boldrini Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy

Department of Telecommunications, Supélec, Gif-sur-Yvette, France

Jakob Lindholm Buthler Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Giuseppe Caso Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy

Andrea Fabio Cattoni Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Symeon Chatzinotas SnT, University of Luxembourg, Walferdange, Luxembourg

Ioannis P. Chochliouros Hellenic Telecommunications Organization S.A. (OTE), Maroussi, Greece

Floriana L. Crespi CSP – ICT innovation, Turin, Italy

Gustavo Wagner Oliveira da Costa Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Luiz A. da Silva Virginia Tech Research Center – Arlington, Arlington, VA, USA Trinity College Dublin, Dublin 2, Ireland

Panagiotis Demestichas University of Piraeus, Piraeus, Greece

Daniel Denkovski Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje, Macedonia

Maria-Gabriella Di Benedetto Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy

Linda Doyle CTVR, Trinity College, Univeristy of Dublin, Dublin, Ireland

Salim Eryigit Department of Computer Engineering, Bogazici University, Istanbul, Turkey

Jocelyn Fiorina Department of Telecommunications, Supélec, Gif-sur-Yvette, France

Tim Forde CTVR, Trinity College, Univeristy of Dublin, Dublin, Ireland

Carolina Fortuna Jožef Stefan Institute, Ljubljana, Slovenia

Liljana Gavrilovska Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje, Macedonia

Andreas Georgakopoulos University of Piraeus, Piraeus, Greece

Fabrizio Granelli DISI – University of Trento, Trento, Italy

Oliver Holland Institute of Telecommunications, King's College London, London, UK

Dimitrios Karvounas University of Piraeus, Piraeus, Greece

Dzmitry Kliazovich University of Luxembourg, Walferdange, Luxembourg

Neumar Malheiros University of Campinas, Campinas, Brazil

Yahia Medjahdi CNAM, Paris, France

João Paulo Miranda CTVR, Trinity College, Univeristy of Dublin, Dublin, Ireland

Mihael Mohorčič Jožef Stefan Institute, Ljubljana, Slovenia

Preben E. Mogensen Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Luca De Nardis Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy

Björn Ottersten SnT, University of Luxembourg, Walferdange, Luxembourg

Ioanna Papafili Hellenic Telecommunications Organization S.A. (OTE), Maroussi, Greece

Alberto Perotti CSP – ICT Innovation, Turin, Italy

Valentin Rakovic Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University, Skopje, Macedonia

Ronald Raulefs Institute of Communications and Navigation of the German Aerospace Center (DLR), Wessling, Germany

Daniel Riviello CSP – ICT Innovation, Turin, Italy

Daniel Roviras CNAM, Paris, France

Didier Le Ruyet CNAM, Paris, France

Shree Krishna Sharma SnT, University of Luxembourg, Walferdange, Luxembourg

Tomaž Šolc Jožef Stefan Institute, Ljubljana, Slovenia

Anastasia S. Spiliopoulou Hellenic Telecommunications Organization S.A. (OTE), Maroussi, Greece

Vera Stavroulaki University of Piraeus, Piraeus, Greece

Paul Sutton CTVR, Trinity College, Univeristy of Dublin, Dublin, Ireland

Michel Terré CNAM, Paris, France

Oscar Tonelli Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Alessandro Tosti Telecom Italia, Rome, Italy

Kostas Tsagkaris University of Piraeus, Piraeus, Greece

Tuna Tugcu Department of Computer Engineering, Bogazici University, Istanbul, Turkey

H. Birkan Yilmaz Department of Computer Engineering, Bogazici University, Istanbul, Turkey

Siwei Zhang Institute of Communications and Navigation of the German Aerospace Center (DLR), Wessling, Germany

Chapter 1
New Types of Air Interface Based on Filter Banks for Spectrum Sharing and Coexistence

Yahia Medjahdi, Didier Le Ruyet, Daniel Roviras, and Michel Terré

Abstract Today, we witness a continuous evolution of applications for wireless communications. In fact, there is a growing interest in the design and the development of cognitive radio technology to overcome the problem of spectrum scarcity resulting from conventional static spectrum allocation. The concept of cognitive radio is based on opportunistic access to the available frequency resources. It offers to future communication systems the ability to dynamically and locally adapt their operating spectrum by selecting it from a wide range of possible frequencies.

Multicarrier techniques are promising and potential candidates offering flexible access to these new spectrum opportunities. Indeed, OFDM, which is the most commonly used multicarrier technique, has been adopted in IEEE 802.22 standard for unlicensed wireless regional area network (WRAN) considering cognitive communications on the unused TV bands. Unfortunately, OFDM presents some weaknesses. In fact, the redundancy, caused by the insertion of the cyclic prefix mandatory part of the transmitted OFDM symbol, reduces the useful data rate. Furthermore, the poor spectral localization of the OFDM subcarriers due to the large sidelobes induces not only an additional spectral loss but also interference problems with unsynchronized signals. These shortcomings have stimulated the research of an alternative scheme that can overcome these problems. An enhanced physical layer based on the filter bank processing called filter bank-based multicarrier (FBMC) technique has been proposed in various works. In fact, a major improvement in spectrum efficiency and better flexibility for system coexistence can be achieved by FBMC thanks to the use of spectrally well-localized waveforms.

Y. Medjahdi (✉) • D.L. Ruyet • D. Roviras • M. Terré
CNAM, 292 rue Saint Martin, 75141 Paris, France
e-mail: yahia.medjahdi@cnam.fr; leruyet@cnam.fr; daniel.roviras@cnam.fr; michel.terre@cnam.fr

The goal of this chapter is to present a tutorial review of the fundamental theory of FBMC techniques. First, we succinctly explain the general concept of multi-carrier transmissions. The advantages and the drawbacks of the classical OFDM are then discussed. Next, we introduce the different schemes of FBMC systems: Filtered MultiTone (FMT), Cosine modulated MultiTone (CMT) and Staggered MultiTone (SMT, also called FBMC-OQAM). Then, we develop theoretical background of the FBMC transmission. Furthermore, the polyphase implementation of the filter bank transceiver and the prototype filter design are reviewed. Finally, we highlight the interest of FBMC spectrum sharing and investigate coexistence issues.

1.1 The Concept of Multicarrier Transmission

Let us consider a single carrier transmission scheme with linear modulation (e.g. M-QAM) and a symbol duration T. Let B the frequency bandwidth which is typically of the order of the inverse of the symbol duration (T^{-1}). For a propagation channel with a delay spread τ_{ds}, we can avoid the intersymbol interference (ISI) if [24]

$$\tau_{ds} \ll T \tag{1.1}$$

On the other hand for $\tau_{ds} > T$, equalization at the receiver can deal with the generated ISI. However when $\tau_{ds} \gg T$, equalization complexity becomes so high that multicarrier modulations (MCM) are preferred.

The key idea in multicarrier transmission to reduce the equalization complexity is to split the data stream into N parallel low-rate streams which are transmitted on N adjacent subcarriers. The bandwidth of each subcarrier is B/N, while the symbol duration T is multiplied by a factor of N. However, the choice of the number of subcarriers N must also provide the robustness of the transmission to the time incoherence of the channel which is related to the maximum Doppler frequency v_{max}. This can be achieved by fulfilling the following condition

$$v_{max}T \ll 1 \tag{1.2}$$

Therefore to obtain the best possible transmission for a given propagation channel characterized by parameters (τ_{ds}, v_{max}), we have to choose the optimal symbol duration that satisfies both conditions at the same time.

1.1.1 Gabor Analysis of Multicarrier Systems

The symbol period T and the subcarrier spacing F are the main parameters characterizing any multicarrier transmission. Theses parameters can be investigated

via Gabor analysis, where the product $T \times F$ stands for the symbol density of a given multicarrier system. It determines in addition to the spectral efficiency of the transmission, the representability of the signal space and the invertibility of the transformation between information symbols and constructed multicarrier signal [46]. The modulation in frequency and the translation in time of a given prototype filter constitute what we call the modulation basis or Gabor family. This latter should be linearly independent to ensure the invertibility of the transformation allowing a perfect reconstruction of the signal. We can distinguish three classes of multicarrier systems with respect to the spectral efficiency of the systems, the linear independency and the completeness of the modulation basis [33, 51]:

1.1.1.1 Undersampled Lattice ($T \times F > 1$)

Gabor family, in this case, cannot be a complete basis since the lattice is not sufficiently sampled. However, linear independency between the basis functions is still possible. Therefore, well-localized prototype filters obtained from incomplete bases can be adapted to the communication systems.

1.1.1.2 Critically Sampled Lattice ($T \times F = 1$)

In this case, we have a complete Gabor system and orthogonal bases exist. However, according to Balian-Low theorem [22], such systems are not able to use well-localized waveforms maintaining in the mean time the orthogonality condition. Therefore, this multicarrier class is using bad localized waveforms like rectangular window filter. In [33], it has been demonstrated that the perfect reconstruction might not be possible beyond the critically sampled grid condition. Hence, the spectral efficiency is maximized for $T \times F = 1$.

1.1.1.3 Oversampled Lattice ($T \times F < 1$)

An overcomplete Gabor family is obtained in this case and it cannot be considered as a basis. However, since Gabor family is overcomplete, representation of a signal might not be unique leading to a lossy reconstruction [51].

1.1.2 Orthogonal and Biorthogonal Multicarrier Systems

A multicarrier system is orthogonal when an orthogonal modulation basis is used at both the transmitter and the receiver. In other words, the transmit filter is equal to the receive filter. This corresponds to the match filtering which maximizes the signal-to-noise ratio (SNR) in the additive white Gaussian noise (AWGN) channel.

Furthermore, perfect reconstruction is still possible by using two different orthogonal Gabor systems at the transmitter and the receiver. In that case, we have a biorthogonal system. In contrast to orthogonal systems, biorthogonal ones are more robust to the dispersiveness of the propagation channel [52].

1.2 Orthogonal Frequency Division Multiplexing (OFDM): Advantages and Drawbacks

OFDM is the most well documented MCM scheme. It has also been adopted by various practical and commercial systems such as, digital audio broadcasting (DAB) [16, 30] terrestrial digital video broadcasting (DVB-T) [17], and the IEEE 802.11a wireless local area network (WLAN) [25], etc. This success comes from a number of advantages that it offers:

- Its robustness to multipath fading effects and its ability to avoid both intersymbol and intercarrier interference (ISI and ICI) by appending a cyclic prefix (CP) that significantly reduces the complexity of the equalization to a single complex coefficient per subcarrier equalizer. This holds as long as the CP covers the maximum delay spread of the channel τ_{ds}.
- The digital implementation of both OFDM modulator and demodulator can be efficiently realized making use of fast Fourier transform (FFT).
- Closely spaced orthogonal subcarriers that divide the available bandwidth into a maximum collection of narrow subchannels.
- Adaptive modulation schemes that can be applied to maximize the spectral efficiency.

Nevertheless, in spite of these advantages, OFDM has some drawbacks:

- There is a loss in spectral efficiency due to the cyclic prefix insertion. Indeed, the cyclic prefix is a copy of part of the transmitted OFDM symbol, and this redundancy reduces the effective throughput.
- OFDM is very sensitive to residual frequency and timing offsets that can be generated by a defective synchronization as well as the Doppler effect [26, 42].
- The use of a rectangular impulse response in OFDM causes large sidelobes at each subcarrier. Hence, the subchannels at the edge of the transmission bandwidth could be a source of interference for other neighboring systems [40, 42].

1.3 Filter Bank Based Multicarrier Systems

To overcome some CP-OFDM shortcomings such as the poor spectrum localization while maintaining some of its advantages, a finite pulse shape filter (or prototype filter) different from the rectangular one with smooth edges can be used [21]. This

leads to the filtered OFDM [61] and the filter-bank based multicarrier (FBMC) systems [9, 11, 47]. On the other hand, the insertion of the cyclic prefix or any guard interval leads to a spectral efficiency loss. Hence, to deal with this limitation, we have to choose a multicarrier scheme with symbol density of one.

However, according to the Balian-Low theorem, see [22], it is theoretically proven with that it is not possible to get a prototype function being well-localized in time and frequency, and satisfying in the meantime the complex orthogonality condition and a symbol density of one.

Therefore, in order to meet these objectives, we must relax the complex orthogonality condition and restrict it to the real field. Thus, to be able to recover the data at the receiver side, the transmitted data must be real-valued (or purely imaginary) chosen from a pulse amplitude modulation (PAM) instead of a quadrature amplitude modulation (QAM). But to maintain the same desired data rate, we should choose a symbol density of two real-valued symbols (to be equivalent to the one complex-valued symbol) per area unit of the time-frequency plane [21].

The pioneering work of Chang [9] in the 1960s, has presented the conditions required for signaling a parallel set of PAM symbols sequences through a bank of overlapping vestigial side-band (VSB) modulated filters. This idea has been extended by Saltzberg in [47], where he showed that the Chang's method could be adapted for the transmission of QAM symbols. Efficient digital implementation of Saltzberg technique through polyphase networks was first introduced by Bellanger [5] and later studied by Hirosaki in [29].

The advancements in digital subscriber line (DSL) technology motivated more activity beyond discrete multitone (DMT) (the equivalent terminology used in DSL literature instead of OFDM) and led to two classes of FBMC systems, namely, filtered multitone (FMT) modulation and discrete wavelet multitone (DWMT) [48], which was later renamed to cosine modulated multitone (CMT) [19].

1.3.1   Classes of FBMC

In the literature, there are mainly three forms of filter bank based multicarrier systems (FBMC) that have been studied. We give here a brief presentation of these FBMC forms:

1.3.1.1   FMT

This form of FBMC has been proposed for filtering the narrowband interference in very-high-speed digital subscriber lines (VDSL) [10, 11] and wireless communications [6]. Similar to frequency division multiplexing methods, in FMT there is no overlapping between adjacent subcarriers leading to the absence of ICI. The perfect reconstruction condition related to the absence of ISI can be ensured by using a square root Nyquist filter, as in conventional single carrier systems.

Compared to conventional CP-OFDM where the symbol duration $T_{\text{OFDM}} = T + \Delta$ (Δ is the CP duration) the subcarrier spacing $F_{\text{OFDM}} = 1/T$, the FMT symbol period is $T_{\text{FMT}} = T$ and the subcarrier spacing $F_{\text{FMT}} = (1 + \alpha)/T$, where α is the roll-off factor that determines the excess bandwidth of each subcarrier. The time-frequency lattice related to FMT is depicted in Fig. 1.1a.

1.3.1.2 CMT

As noted earlier, the CMT system is fundamentally the same as DWMT one where parallel streams of PAM data symbols are transmitted using a set of VSB subcarrier channels [21]. This scheme also has the maximum possible bandwidth efficiency. In a CMT system, in order to transmit N complex symbols on each multicarrier symbol, a system with $2N$ subcarrier is implemented where each carrier conveys a real symbol, while, in a FBMC-OQAM system the transceiver would have N subcarriers that convey N complex symbols. If FBMC-OQAM symbols are transmitted at the rate of $1/T$ complex symbols on each subcarrier with a bandwidth of $1/T$, an equivalent CMT system with the same data rate, would have a rate of $1/T$ real symbols on each subcarrier with a bandwidth of $1/2T$. Therefore, the same bandwidth is divided into twice as many subcarriers in case of CMT to achieve the same data rate (see Fig. 1.1b).

1.3.1.3 FBMC-OQAM (SMT, Staggered Multitone [21])

Offset quadrature amplitude modulation (OQAM) is a modified version of QAM. Saltzberg in [47] showed that by choosing a root-Nyquist filter with symmetric impulse response for pulse-shaping at the transmitter and using the same as a match filter at the receiver in a multichannel QAM system, and by introducing a half symbol space delay between the in-phase and quadrature components of the QAM symbols, it is possible to achieve baud-rate spacing between adjacent subcarrier channels and still recover the information symbols, free of both ISI and ICI. This method provides higher spectral efficiency compared to conventional OFDM due to the non-use of any cyclic prefix.

In this chapter, we mainly focus on FBMC-OQAM scheme where we use the terminology "filter bank based multicarrier (FBMC)" to refer to FBMC system that uses OQAM modulation.

1.3.2 Filter Bank Transceivers

The FBMC transmitter system is depicted in Fig. 1.2a. The high-rate data stream is split into N parallel low-rate data streams and transmitted on N subcarriers that are closely spaced at a regular frequency spacing $\Delta f = 1/T$ (T is the
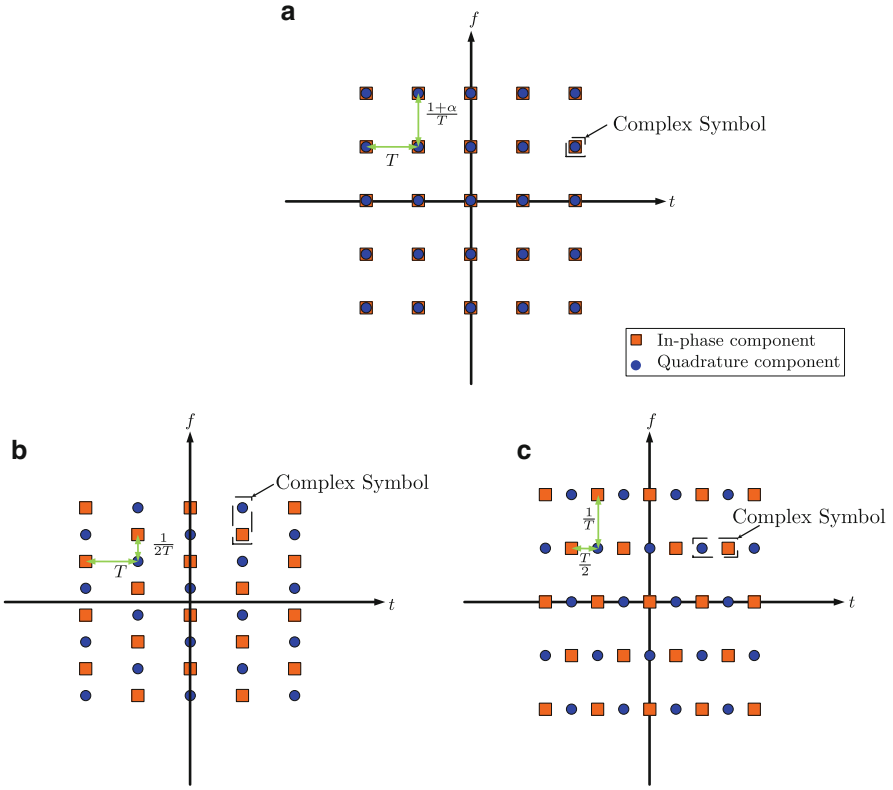
Fig. 1.1 Time-frequency phase-space lattice representation of: (**a**) FMT, (**b**) CMT and (**c**) SMT systems [21]

symbol period). The main idea in this scheme is to transmit OQAM data symbols instead of conventional QAM ones [7, 50], where the in-phase and the quadrature components are time staggered by half a symbol period, $T/2$. The second specificity of this scheme is that considering two successive subcarriers, the time delay $T/2$ is introduced into the imaginary part of the QAM symbols on one of the subcarriers, whereas it is introduced into the real part of the symbols on the adjacent one as shown in Fig. 1.1c.

Suppose that we have the following complex input symbols,

$$x_{m,n} = x_{m,n}^I + j x_{m,n}^Q \qquad (1.3)$$

where $x_{m,n}^I$ and $x_{m,n}^Q$ are the real and the imaginary parts of the n-th symbol of the m-th subcarrier, respectively. The continuous-time baseband transmitted signal is defined as,

Fig. 1.2 The structure of the continuous time FBMC: (**a**) transmitter (**b**) receiver
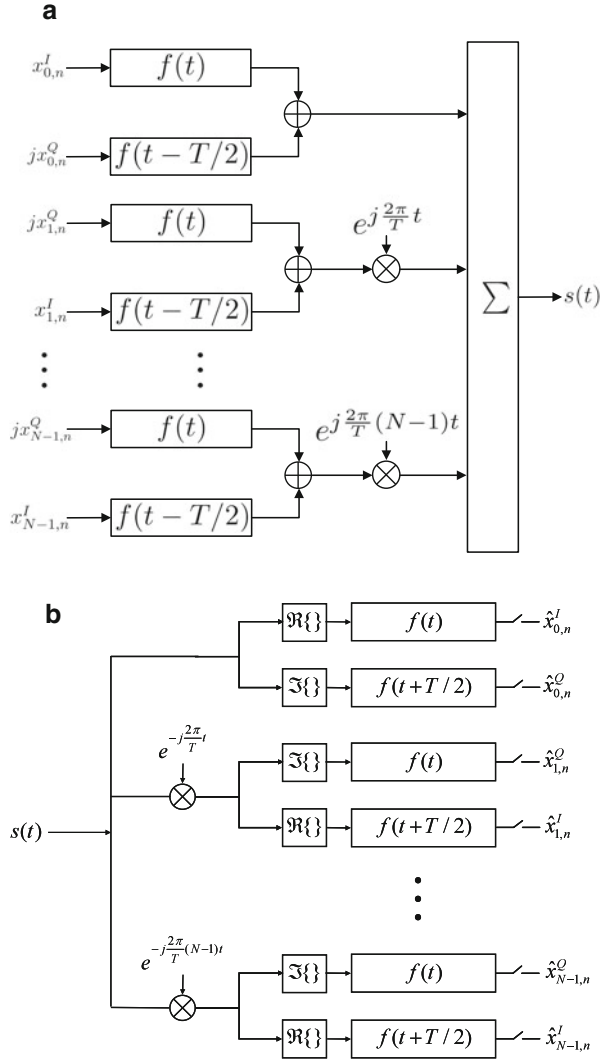
$$s(t) = \sum_{m=0}^{M-1} \sum_{n=-\infty}^{+\infty} (x_{2m,n}^I f(t - nT) + jx_{2m,n}^Q f(t - nT - T/2))e^{j\frac{2\pi}{T}(2m)t}$$

$$+ (x_{2m+1,n}^I f(t - nT - T/2) + jx_{2m+1,n}^Q f(t - nT))e^{j\frac{2\pi}{T}(2m+1)t}$$

$$(1.4)$$

where $f(t)$ is the impulse response of the prototype filter and the number of subcarriers $N = 2M$.

Let us introduce the following notations,

$$a_{2m,2n} = x_{2m,n}^I, \qquad a_{2m,2n+1} = x_{2m,n}^Q$$

$$a_{2m+1,2n} = x_{2m+1,n}^Q, \qquad a_{2m+1,2n+1} = x_{2m+1,n}^I \qquad (1.5)$$

Hence, we rewrite the transmitted signal $s(t)$ as follows,

$$s(t) = \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} f(t - nT/2) e^{j\frac{2\pi}{T}mt} e^{j\varphi_{m,n}}$$

$$= \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} \gamma_{m,n}(t) \qquad (1.6)$$

where

$$\varphi_{m,n} = \frac{\pi}{2}(m+n) - \pi mn \qquad (1.7)$$

and

$$\gamma_{m,n}(t) = f(t - nT/2) e^{j\frac{2\pi}{T}mt} e^{j\varphi_{m,n}} \qquad (1.8)$$

Assuming an ideal transmission system, the output of the receiver (see Fig. 1.2b), $\hat{x}_{m,n}$, consists of the real and the imaginary parts $\hat{x}_{m,n}^I$, $\hat{x}_{m,n}^Q$.

$$\hat{x}_{m,n} = \hat{x}_{m,n}^I + j\hat{x}_{m,n}^Q \qquad (1.9)$$

We also denote,

$$\hat{a}_{2m,2n} = \hat{x}_{2m,n}^I \qquad \hat{a}_{2m,2n+1} = \hat{x}_{2m,n}^Q$$

$$\hat{a}_{2m+1,2n} = \hat{x}_{2m+1,n}^Q \qquad \hat{a}_{2m+1,2n+1} = \hat{x}_{2m+1,n}^I \qquad (1.10)$$

Therefore, $\hat{a}_{m,n}$ can be found as the real part of the signal at the output of the corresponding receive filter with response $f(t)$ and expressed as,

$$\hat{a}_{m,n} = \Re\left[\langle s(t), \gamma_{m,n}(t)\rangle\right] = \Re\left[\int_{-\infty}^{+\infty} \gamma_{m,n}^*(t)s(t)dt\right] \qquad (1.11)$$

Substituting (1.6) and (1.8) in (1.11), we obtain

$$\hat{a}_{m,n} = \sum_{m'=0}^{N-1} \sum_{n'=-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Re\left[a_{m',n'} f(t - nT/2) f(t - n'T/2) e^{j\frac{2\pi}{T}(m'-m)t} e^{j(\varphi_{m',n'}-\varphi_{m,n})} dt\right]$$

$$(1.12)$$

We now proceed to discuss the design of the receive filter $f(t)$. In a distortion-free channel, the received signal equals the transmitted one

$$\hat{a}_{m,n} = a_{m,n} \qquad (1.13)$$

For convenience of the design, it is common to constrain $f(t)$ to be real and symmetric (i.e. $f(t) = f(-t)$). Therefore, (1.13) can be met, if $f(t)$ is chosen such that the following identity is satisfied,

$$\int_{-\infty}^{+\infty} f(t - nT/2) f(t - n'T/2) \cos\left(\frac{2\pi}{T}(m' - m)t + \varphi_{m',n'} - \varphi_{m,n}\right) dt = \delta_{m,m'} \delta_{n,n'}$$

$$(1.14)$$

The signal $s(t)$ propagates through a frequency selective multipath channel, where its equivalent sample-spaced impulse response is given by

$$h(t) = \sum_{i=0}^{L-1} h_i \delta\left(t - \frac{n_i}{N}T\right) \qquad (1.15)$$

where $n_0 < n_1 < \ldots < n_{L-1} < C$ and C is the maximum delay spread of the channel normalized by the sampling period (T/N), and h_i are the complex channel path gains, which are assumed mutually independent, where $\mathbb{E}[h_i h_i^*] = \gamma_i$, and $\mathbb{E}[h_i h_j^*] = 0$ when $i \neq j$. We further assume that the power is normalized such that $\sum_{i=0}^{L-1} \gamma_i = 1$.

We also assume that the propagation channel is stationary over one OFDM symbol. This is the case for time-invariant or slowly time-varying channels. This assumption is valid in most of wireless applications. In these application, the channel coherence time is much larger than the block duration and the assumption of a static channel is reasonable [43].

The baseband version of the received signal can thus be written in continuous time as follows,

$$r(t) = h(t) \star s(t) \qquad (1.16)$$

Substituting (1.6) and (1.15) in (1.16), we obtain

$$r(t) = \sum_{i=0}^{L-1} h_i \delta\left(t - \frac{n_i}{N}T\right) \star \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} f(t - nT/2) e^{j\frac{2\pi}{T}mt} e^{j\varphi_{m,n}}$$

$$= \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j\varphi_{m,n}} \sum_{i=0}^{L-1} h_i f\left(t - nT/2 - \frac{n_i}{N}T\right) e^{j\frac{2\pi}{T}m(t - \frac{n_i}{N}T)} \qquad (1.17)$$

Without loss of generality, let KT be the duration of the prototype filter $f(t)$, where K is an integer greater or equal to 1 (i.e. $K \geq 1$). Thus, we can say that the bandwidth occupied by the filter is smaller than the coherence bandwidth of the channel $B_c = 1/(2\tau_{ds})$ [24, 53]. We recall that τ_{ds} denotes the maximum delay spread of the channel.

Looking at (1.17), we can notice that $f(t - nT/2 - \tau)$ may have relatively slow variations when $\tau \in [0, \tau_{ds}]$ [36, 37]. Indeed, compared to the coherence bandwidth B_c, the filter bandwidth is very small, which also means that the time variations of the prototype filter $f(t)$ are necessarily limited.

Consequently, (1.17) becomes,

$$
r(t) = \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j\varphi_{m,n}} f(t - nT/2) e^{j\frac{2\pi}{T}mt} \sum_{i=0}^{L-1} h_i e^{-j\frac{2\pi}{N}mn_i}
$$

$$
= \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j\varphi_{m,n}} f(t - nT/2) e^{j\frac{2\pi}{T}mt} H(m) \tag{1.18}
$$

where, $H(m) = \sum_{i=0}^{L-1} h_i e^{-j\frac{2\pi}{N}mn_i}$ is the complex channel gain at subcarrier m.

The demodulated signal y_{m_0,n_0} at time instant n_0 and subcarrier m_0 is given by

$$
y_{m_0,n_0} = \int_{-\infty}^{+\infty} \gamma_{m,n}^*(t) r(t) dt
$$

$$
= \sum_{m=0}^{N-1} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j(\varphi_{m,n} - \varphi_{m_0,n_0})} H(m) \int_{-\infty}^{+\infty} f(t - nT/2) f(t - n_0 T/2) e^{j\frac{2\pi}{T}(m-m_0)t} dt \tag{1.19}
$$

Now, let $\Omega_{\Delta m}$ be the neighborhood area around the subchannel m_0 where the channel can be considered to be constant,

$$
\Omega_{\Delta m} = \{l, |l| \leq \Delta m \mid H(m_0 + l) \approx H(m_0)\}
$$

It should be noticed that $\Omega_{\Delta m}$ depends on the coherence bandwidth B_c, i.e., on τ_{ds} [24, 53].

If we consider that the prototype filter is well localized in both time and frequency domains [35, 50] meaning that, $\int_{-\infty}^{+\infty} f(t - nT/2) f(t - n_0 T/2) e^{j\frac{2\pi}{T}(m-m_0)t} dt$ immediately tends to zero when $|n - n_0|$ and $|m - m_0|$ increase, we can write

$$
y_{m_0,n_0} = H(m_0) \sum_{l \in \Omega_{\Delta m}} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j(\varphi_{m_0+l,n} - \varphi_{m_0,n_0})} \int_{-\infty}^{+\infty} f(t - nT/2) f(t - n_0 T/2) e^{j\frac{2\pi}{T}lt} dt \tag{1.20}
$$

Assuming a perfect channel equalization, we obtain the resulting signal y_{m_0,n_0}^{eq}

$$y_{m_0,n_0}^{eq} = \sum_{l\in\Omega_{\Delta m}} \sum_{n=-\infty}^{+\infty} a_{m,n} e^{j(\varphi_{m_0+l,n}-\varphi_{m_0,n_0})} \int_{-\infty}^{+\infty} f(t-nT/2)f(t-n_0T/2)e^{j\frac{2\pi}{T}lt}\,dt$$

(1.21)

After the OQAM decision, we write the recovered symbol $\hat{a}_{m_0,n_0}$ as follows,

$$\hat{a}_{m_0,n_0} = \Re\left\{y_{m_0,n_0}^{eq}\right\}$$

$$= \sum_{l\in\Omega_{\Delta m}} \sum_{n=-\infty}^{+\infty} a_{m,n} \int_{-\infty}^{+\infty} f(t-nT/2)f(t-n_0T/2)\cos\left(\frac{2\pi}{T}lt+\varphi_{m_0+l,n}-\varphi_{m_0,n_0}\right)dt$$

(1.22)

According to the orthogonality condition of the prototype filter given by (1.14), we obtain

$$\hat{a}_{m_0,n_0} = \sum_{l\in\Omega_{\Delta m}} \sum_{n=-\infty}^{+\infty} a_{m,n}\delta_{m_0+l,m_0}\delta_{n,n_0} = a_{m_0,n_0}$$

(1.23)

In the case of highly frequency selective channels, the assumption that $f(t-\tau) \approx f(t)$ when $\tau \in [0, \tau_{\text{ds}}]$ is no longer valid. In the following paragraph, we give a brief review of the different techniques of equalization of frequency selective channels.

1.3.3 Equalization Techniques in FBMC

Three main approaches have been proposed in the literature to deal with frequency selective channels. The first one uses well localized waveforms that is, the pulse energy both in time and frequency domains are well contained to limit the effect on the neighborhood of a given symbol [35, 50, 54]. In this case, a basic equalizer structure of a single complex coefficient per subcarrier is considered.

The second approach uses FIR (finite impulse response) filters as subcarrier equalizers with cross connections between the adjacent subchannels to cancel the ICI [48, 57].

The third approach applies a receiver filter bank structure providing over-sampled subcarrier signals to avoid the cross connections between the subchannels, and performs subcarrier equalization using FIR filters [28, 44, 56]. Recently, Waldhauser et al. have proposed MMSE (minimum mean square error) and decision feedback equalizer per subcarrier designed for FBMC/OQAM [59, 60]. Based on the same

approach but using frequency sampling method, Ihalainen et al. have presented a multi-tap per-subcarrier equalizer in such a manner that the frequency response of the designed filter is forced to take the given target values at a set of considered frequency points within a subchannel [31]. Another interesting work has proposed a scheme of equalization with interference cancelation [37]. The basic principle of this method rests on an accurate computation of the interference term taking into account the channel and the prototype filter coefficients.

1.3.3.1 Review of Multiple Taps Equalizers Design

The derivation of the equalizer coefficients is based on the principle that the equalizer of a subcarrier is designed to optimally compensate, at some points in the sub-band, the channel distortions and the timing offset between the transmitter and the receiver. More specifically, the equalizer coefficients are computed such that, the equalizer amplitude response perfectly meets the inverse of the channel amplitude response and the phase equalizer is equal to the negative phase of the channel response, at all considered frequency points.

In this section, we discuss the design of the 3-tap equalizer. It is worth noticing that any n-multi-tap equalizer can be computed in the same manner. In this case, three points of the inverse channel frequency response are required for each subchannel: $EQ(i)$ and two intermediate points $EQ1$, $EQ2$ (as depicted in Fig. 1.3a).

As previously mentioned, the frequency points $EQ(i)$, $EQ1$ and $EQ2$ for a given subchannel i are computed following the ZF (Zero-Forcing) criterion, that is,

$$EQ(i) = \frac{H^*(e^{j2\pi i/N})}{|H(e^{j2\pi i/N})|^2}$$

$$EQ1 = \frac{H^*(e^{j\pi(2i-1)/N})}{|H(e^{j\pi(2i-1)/N})|^2}$$

$$EQ2 = \frac{H^*(e^{j\pi(2i+1)/N})}{|H(e^{j\pi(2i+1)/N})|^2} \tag{1.24}$$

where,

- $H(e^{j2\pi i/N})$ is the channel frequency response at the subcarrier i
- H^* stands for the conjugate of H

Let $C_{\text{eq,i}}(z)$ be the response of the equalizer of the subchannel m,

$$C_{\text{eq,i}}(z) = c_{-1,i}z + c_{0,i} + c_{+1,i}z^{-1} \tag{1.25}$$
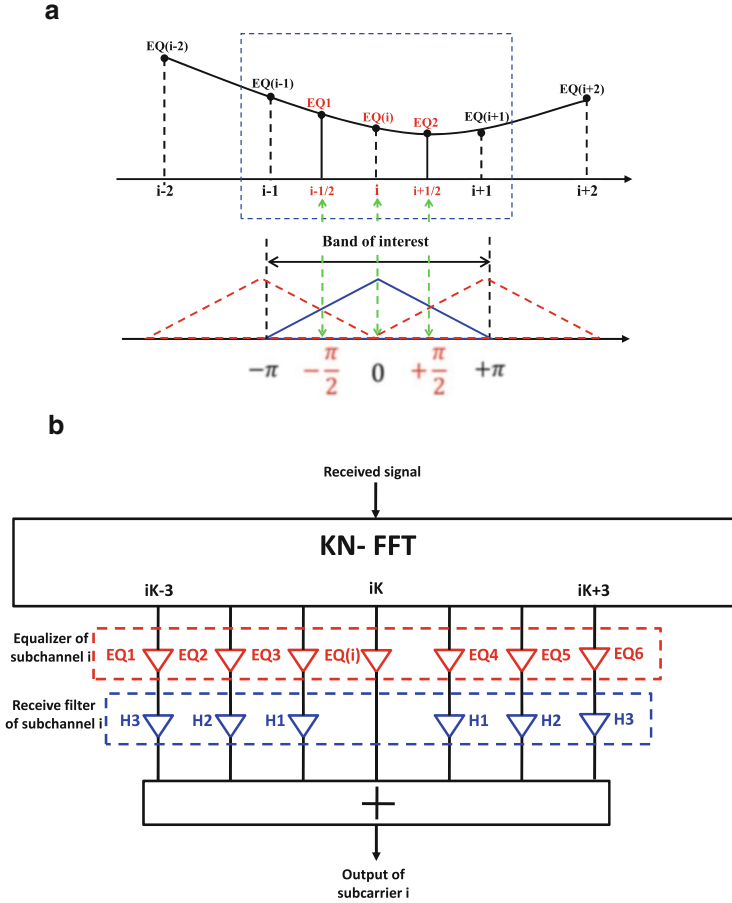
Fig. 1.3 (**a**) Points of the inverse subchannel frequency response used to compute the multi-tap equalizer, (**b**) frequency domain equalizer for subchannel i in FS-FBMC scheme [3]

According to [31], the equalizer coefficients can be computed by resolving the following system of equations,

$$
\begin{cases}
C_{\mathrm{eq},i}(e^{-j\pi/2}) = EQ1 \\
C_{\mathrm{eq},i}(e^{-j0}) = EQ(i) \quad \text{(even subchannels) or} \\
C_{\mathrm{eq},i}(e^{+j\pi/2}) = EQ2
\end{cases}
$$

$$
\begin{cases}
C_{\mathrm{eq},i}(e^{j\pi/2}) = EQ1 \\
C_{\mathrm{eq},i}(e^{j\pi}) = EQ(i) \quad \text{(odd subchannels)} \\
C_{\mathrm{eq},i}(e^{j3\pi/2}) = EQ2
\end{cases}
\tag{1.26}
$$

We then obtain,

$$c_{-1,i} = \pm \frac{1}{4} \left[(2EQ(i) - EQ1 - EQ2) - j(EQ2 - EQ1) \right]$$

$$c_{0,i} = \frac{1}{2} (EQ1 + EQ2)$$

$$c_{+1,i} = \pm \frac{1}{4} \left[(2EQ(i) - EQ1 - EQ2) + j(EQ2 - EQ1) \right] \qquad (1.27)$$

Note that the signs $+$ and $-$ correspond respectively to even and odd subchannels.

Recently, Bellanger has proposed an alternative scheme for FBMC called frequency spreading filter bank multicarrier (FS-FBMC). In such a scheme, the FBMC receiver consists of an extended FFT with a size $K \times N$ (which is equal to the prototype filter length), followed by a summation of the $2K - 1$ FFT outputs weighted by $2K - 1$ frequency components of the considered prototype filter (e.g. $H_3, H_2, H_1, 1, H_1, H_2, H_3$ for PHYDYAS[1] prototype filter with an overlapping factor $K = 4$), for each subcarrier [3]. It is worth noticing that in this case, the prototype filter is designed based on the frequency sampling technique [38].

In FS-FBMC, the multiple taps equalization is performed in the frequency domain by correcting all the spectral components of the signal as depicted in Fig. 1.3b. Another interesting characteristic of this equalization method is that no additional delay is required, in contrast to time-domain subchannel equalization [3]. However, the counterpart of this scheme is an increase in the computational complexity because an FFT of $K \times N$ should be computed every symbol period.

1.3.4 Polyphase Implementation of Filter Bank for Multicarrier Transmission

FBMC systems are implemented using an efficient technique that is called polyphase implementation. The polyphase representation, proposed by Bellanger over 35 years ago [4], is a pivotal tool in multirate digital signal processing. This form provides great simplification of theoretical analysis and computationally efficient implementations for several filtering applications. The concept of this implementation is outlined in this section.

---

[1]PHYDYAS: EC funded Physical Layer for Dynamic Spectrum Access and Cognitive Radio, http://www.ict-phydyas.org/

Fig. 1.4 The polyphase implementation of the filter $F[z]$
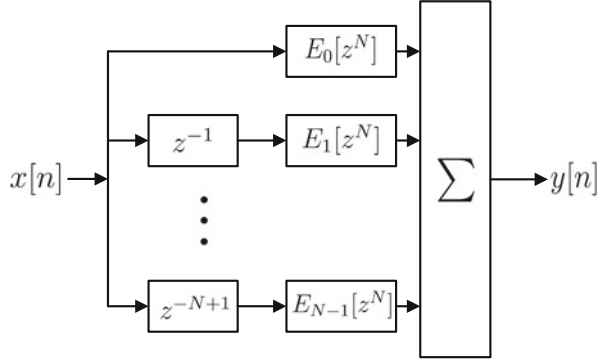
1.3.4.1 Polyphase Representation

Consider a FIR filter $F(z)$ of $L = KN$ coefficients,

$$F(z) = \sum_{l=0}^{L-1} f[l]z^{-l} \tag{1.28}$$

This filter can be decomposed into N elementary filters, in the following way,

$$F(z) = \sum_{l=0}^{L-1} f[l]z^{-l} = \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} f[kN+n]z^{-(kN+n)}$$

$$= \sum_{n=0}^{N-1} \left[\sum_{k=0}^{K-1} f[kN+n]z^{-kN} \right] z^{-n} \tag{1.29}$$

Consequently, the polyphase network (PPN) of the filter $F(z)$ (see Fig. 1.4) is given by,

$$F(z) = \sum_{n=0}^{N-1} E_n[z^N]z^{-n} \tag{1.30}$$

where $E_n[z^N] = \sum_{k=0}^{K-1} f[kN+n]z^{-kN}$ are called the polyphase components.

Now, let $F_i(z)$ be a replica of the filter $F(z)$ shifted by i/N in the frequency domain,

$$F_i(z) = \sum_{l=0}^{L-1} f[l]e^{j\frac{2\pi}{N}il}z^{-l} \tag{1.31}$$

Using the polyphase representation, $F_i(z)$ becomes,

$$F_i(z) = \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} f[kN + n] e^{j\frac{2\pi}{N} i(kN+n)} z^{-(KN+n)}$$

$$= \sum_{n=0}^{N-1} e^{j\frac{2\pi}{N} ni} E_n[z^N] z^{-n} \tag{1.32}$$

A uniform filter bank is obtained by shifting the response of $F(z)$ which is called the "*prototype filter*" on the frequency axis. Considering all shifts multiples of $1/N$: $[1, \ldots, N-1]/N$, we obtain

$$\begin{bmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{N-1}(z) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w^{-1} & \cdots & w^{-(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{-(N-1)} & \cdots & w^{-(N-1)^2} \end{bmatrix}}_{IDFT_N} \underbrace{\begin{bmatrix} E_0[z^N] \\ z^{-1} E_1[z^N] \\ \vdots \\ z^{-(N-1)} E_{N-1}[z^N] \end{bmatrix}}_{PPN} \tag{1.33}$$

where $w = e^{-j\frac{2\pi}{N}}$ and the square matrix is the inverse discrete Fourier transform (IDFT) matrix of order N.

1.3.4.2 Multirate Identities

Before discussing the digital implementation of the FBMC transceiver, it is worth recalling the multirate identities also called "noble identities" [20, 55]. In fact, using these identities the filtering operation can be performed at the lower rate with decimated filters, making the implementation more efficient.

First identity: The up-sampled filtering $E[z^N]$ followed by down-sampling ($\downarrow N$) is equivalent to down-sampling ($\downarrow N$) followed by filtering $E[z]$.

Second identity: Up-sampling ($\uparrow N$) followed by up-sampled filtering $E[z^N]$ is equivalent to filtering $E[z]$ followed by up-sampling ($\uparrow N$).

1.3.4.3 Polyphase Synthesis Filter Bank

The digital implementation of a synthesis filter bank (SFB) is presented in Fig. 1.5a. The output signal $y[n]$ is synthesized by the transmission of a set of parallel data symbols $x_k[n]$ through a bank of modulated filters. According to (1.33), one can design the structure presented in Fig. 1.5b. Since IDFT is a constant matrix
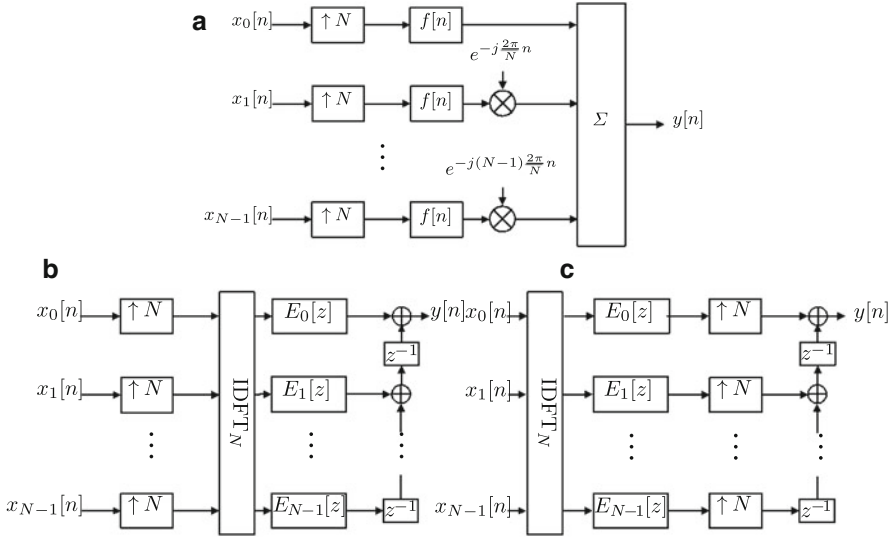
Fig. 1.5 Synthesis filter bank: (**a**) general structure, (**b**) polyphase implementation and (**c**) simplified polyphase implementation

multiplication, it can be switched with the up-sampling. Moreover, according to the second noble identity, we can commute the up-sampling with the polyphase components by replacing $E_n[z^N]$ with $E_n[z]$ and consequently the polyphase implementation of the synthesis filter bank simplifies to the scheme depicted in Fig. 1.5c.

1.3.4.4 Polyphase Analysis Filter Bank

The analysis filter bank (AFB), as indicated by its name, performs a frequency decomposition of the input signal. The polyphase implementation of the AFB represents a structure which is the dual of the previously considered synthesis filter bank, and is derived in a similar manner. The starting point of this is the general structure of N-channel analysis filter bank shown in Fig. 1.6a.

Consider $F_{-i}(z)$ the demodulated version of $F(z)$ (a replica which is shifted in frequency by $-i/N$),

$$F_{-i}(z) = \sum_{l=0}^{L-1} f[l]e^{-j\frac{2\pi}{N}il}z^{-l} \tag{1.34}$$
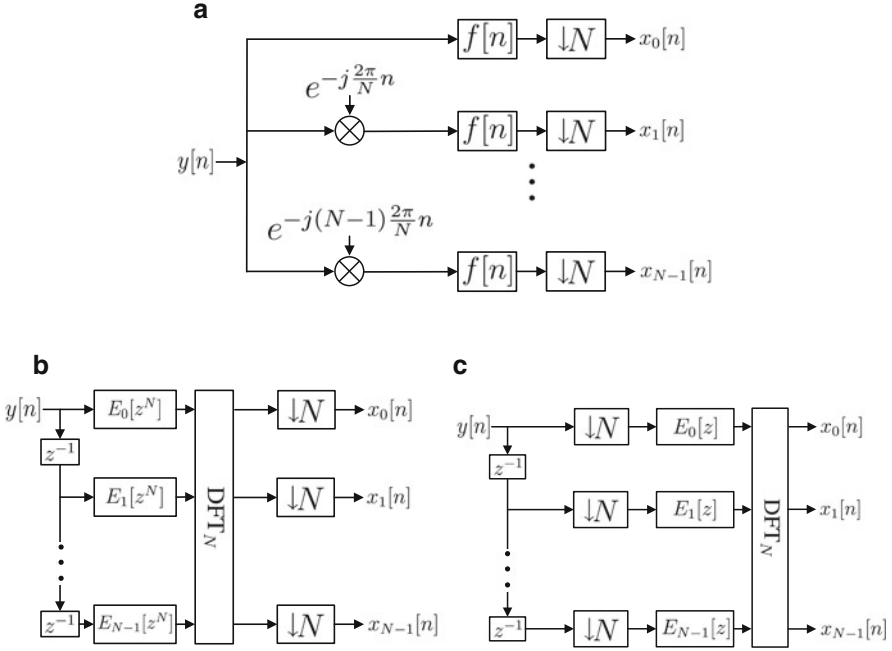
Fig. 1.6 Analysis filter bank: (**a**) general structure, (**b**) polyphase implementation and (**c**) simplified polyphase implementation

Using the polyphase representation, $F_{-i}(z)$ becomes,

$$
F_{-i}(z) = \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} f[kN+n] e^{-j\frac{2\pi}{N} i(kN+n)} z^{-(KN+n)}
$$

$$
= \sum_{n=0}^{N-1} e^{-j\frac{2\pi}{N} ni} E_n[z^N] z^{-n} \tag{1.35}
$$

Thus, we can write

$$
\begin{bmatrix} F_0(z) \\ F_{-1}(z) \\ \vdots \\ F_{-(N-1)}(z) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & w & \cdots & w^{(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{(N-1)} & \cdots & w^{(N-1)^2} \end{bmatrix}}_{DFT_N} \underbrace{\begin{bmatrix} E_0[z^N] \\ z^{-1} E_1[z^N] \\ \vdots \\ z^{-(N-1)} E_N[z^N] \end{bmatrix}}_{PPN} \tag{1.36}
$$

where $w = e^{-j\frac{2\pi}{N}}$ and the square matrix is the discrete Fourier transform matrix of order N.
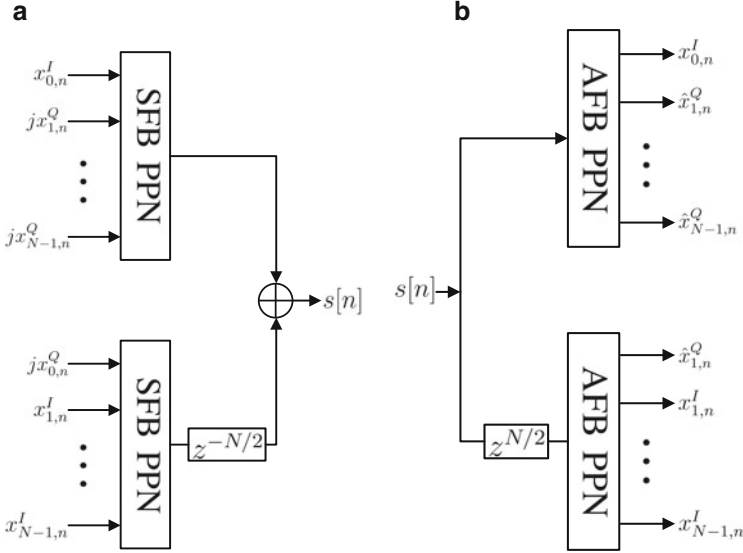
Fig. 1.7 The block diagram of the polyphase implementation of FBMC: (**a**) transmitter, (**b**) receiver

Therefore, the polyphase implementation of the analysis filter bank is depicted in Fig. 1.6b, where the incoming signal is passed through a delay chain, and filtered by the polyphase components $E_n[z^N]$. This filtering operation is followed by DFT operation. Similarly to the synthesis filter banks, the DFT can be commutated with down-sampling and using the first noble identity the polyphase components are switched with down-sampling as presented in Fig. 1.6c.

1.3.4.5   FBMC Polyphase Structure

The structure of the FBMC transmitter filter bank shown in Fig. 1.2a can be split into two subsystems: the first one with prototype filter $f(t)$ and the second one with prototype filter $f(t - T/2)$.

It is worth mentioning that we can move the delay of $N/2$ (which is equivalent to half a symbol period, T/2) from $f(n - N/2)$ to outside the second subsystem. Therefore, by replacing each subsystem with its polyphase implementation, we obtain the FBMC transmitter polyphase structure (see Fig. 1.7a).

Similarly to the modifications performed on the FBMC transmitter, we can rearrange the FBMC receiver structure presented in Fig. 1.2b into two subsystems with prototype filters $f(t)$ and $f(t + T/2)$. Also, the delay $-T/2$ can be displaced to the input of the second subsystem. Doing these changes and using the polyphase implementation of the analysis filter bank, the polyphase structure of the FBMC receiver can be obtained in Fig. 1.7b. This type of structures are used in [23] and [20].

It should be pointed out that other polyphase implementations of the FBMC system has been proposed in the literature e.g. [8, 29, 50] and [58].

1.3.5   Review of Prototype Filter Design

The study of prototype filters is of a particular interest for FBMC systems because it offers an important degree of freedom compared to what is possible with conventional OFDM. A prototype filter has to meet orthogonality constraint that is given in continuous-time for the prototype function by (1.14) or can be related to the polyphase components of the prototype filter in discrete-time as shown in [50]. Furthermore these prototypes can be built to satisfy some target objective, e.g. frequency selectivity or time-frequency localization, regularity, etc. [49].

A basic constraint of data transmission is that the filter must satisfy the Nyquist criterion, to avoid ISI. If the symbol period is T_{symb} and the symbol rate is $f_{symb} = 1/T_{symb}$, the channel frequency response must be symmetrical around the frequency $f_{symb}/2$. Accordingly, in FBMC, the prototype filter for the synthesis and analysis filter banks must be half-Nyquist, which means that the square of its frequency response must satisfy the Nyquist criterion.

So-called perfect-reconstruction (PR) filter banks implement the Nyquist criterion exactly, and also without introducing any cross-talk between subchannels in the back-to-back connection of SFB and AFB (so-called transmultiplexer). In wireless communications, the transmission channel introduces inevitably some distortion to the received subchannel signals. Therefore, the PR condition is not essential, and it is sufficient that the cross-talk between subchannels is small enough to be ignored in comparison to the residual interference, e.g., due to imperfect channel equalization. From the filter bank design point of view, this means that the so-called nearly perfect-reconstruction (NPR) designs are sufficient. Since NPR designs are more efficient than PR designs, e.g., in providing higher frequency selectivity with given prototype filter length, NPR filters are the favored choice in various applications [13].

In the following, we give a brief overview of some examples of prototype functions. The time and frequency representation are given in normalized scales, i.e., by the ratio $\frac{t}{T}$ and $\frac{f}{F}$.

1.3.5.1   Rectangular Window Function

The impulse response of the Rectangular Window (RW) function is defined by

$$f_\Pi(t) = \begin{cases} \frac{1}{\sqrt{T}} & t \in [0, T] \\ 0 & elsewhere \end{cases} \tag{1.37}$$

The frequency response of this filter is given by

$$F_\Pi(f) = \sqrt{T} \times sinc(\pi f T) \tag{1.38}$$

where, $sinc(x) = \sin(x)/x$

It is the same function as the one used in the conventional OFDM. Its time and frequency representations are depicted (*solid line*) in Fig. 1.8a, b, respectively.

1.3.5.2 The Square Root of Raised Cosine (SRRC)

The SRRC pulse response is defined by [12, 45]

$$f_{srrc}(t) = \frac{\sin\left[\pi t(1-\beta)/T\right] + 4\beta t \cos\left[(\pi t/T)(1+\beta)/T\right]/T}{\pi t \left[1 - (4\beta t/T)^2\right]} \tag{1.39}$$

Here, β is called the *roll-off factor* and takes values in the range $0 \le \beta \le 1$.

The frequency response of the SRRC filter is given by [12, 45]

$$F_{srrc}(f) = \begin{cases} \sqrt{T} & |f| \in [0, \frac{1-\beta}{2T}] \\ \sqrt{\frac{T}{2}} \left\{1 + \cos\left[\frac{\pi T}{\beta}\left(|f| - \frac{1-\beta}{2T}\right)\right]\right\}^{1/2} & |f| \in [\frac{1-\beta}{2T}, \frac{1+\beta}{2T}] \\ 0 & |f| > \frac{1+\beta}{2T} \end{cases} \tag{1.40}$$

The time and the frequency responses of the SRRC filter are plotted (dash-dot line with plus sign marker for $\beta = 1$ and dashed line with circle marker for $\beta = 0.5$) in Fig. 1.8a, b, respectively. Comparing the SRRC responses to the rectangular ones, one can see that the SRRC prototype filter presents a significant improvement in terms of frequency localization.

1.3.5.3 Extended Gaussian Function (EGF)

This class of prototype filter results from a two-step orthogonalization algorithm that is applied on the Gaussian function [1, 35]. The EGF is defined in time domain [49] by,

$$f_{\alpha,v_0,\tau_0}(t) = \frac{1}{2}\sum_{k=0}^{\infty} d_{k,\alpha,v_0}\left[g_\alpha\left(t + \frac{k}{v_0}\right) + g_\alpha\left(t - \frac{k}{v_0}\right)\right]\sum_{l=0}^{\infty} d_{l,1/\alpha,\tau_0} \cos\left(2\pi l\frac{t}{\tau_0}\right) \tag{1.41}$$

where, $g_\alpha(t) = (2\alpha)^{1/4}e^{-\pi\alpha t^2}$ and α is the spreading factor. The computation of the real coefficients d_{k,α,v_0} is detailed in [49].
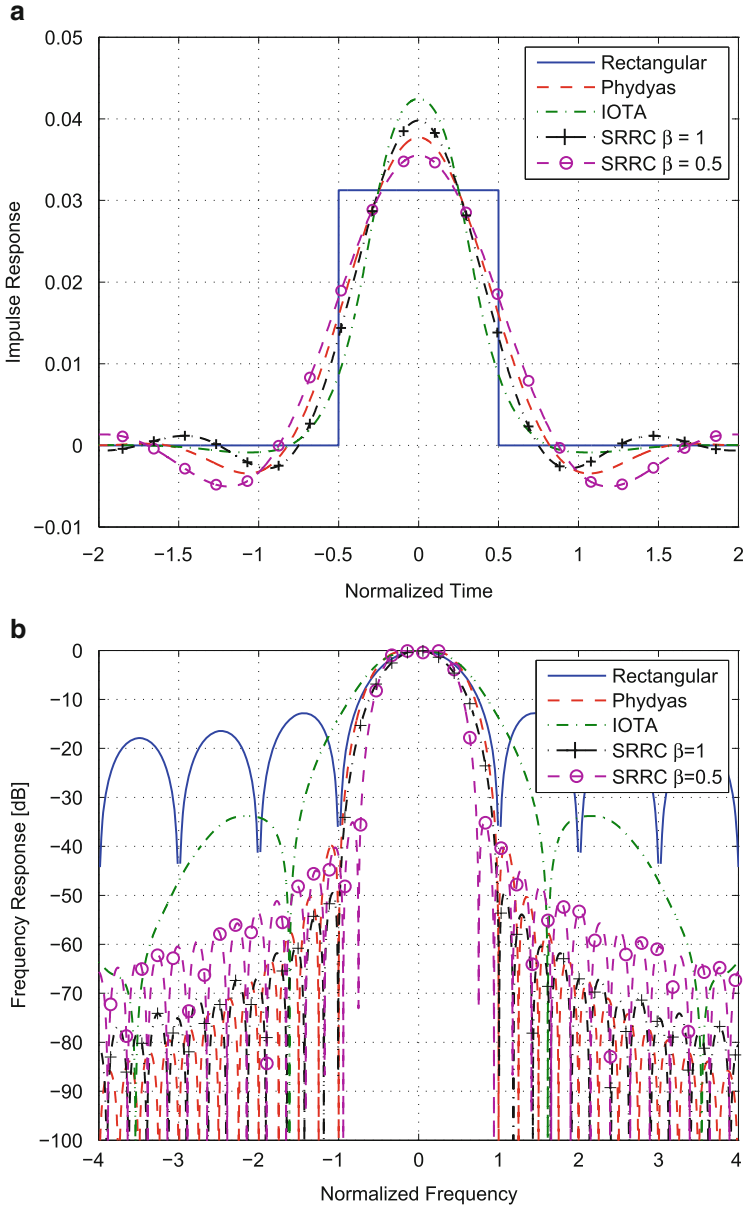
Fig. 1.8 Rectangular window, SRRC, IOTA and PHYDYAS filters: (**a**) impulse responses and (**b**) frequency responses

Originally the EGFs have been derived in closed form by Le Floch et al. [35] using the Isotropic Orthogonal Transform Algorithm (IOTA). Therefore, the IOTA prototype filter is obtained by setting $v_0 = \tau_0 = 1/\sqrt{2}$ and $\alpha = 1$.

Note that the IOTA prototype function is identical to its Fourier Transform [35]. Accordingly, IOTA is equally localized in time and frequency domains. Its excellent time localization implies that it is less sensitive than the SRRC to time truncation [50]. The IOTA time and frequency responses are displayed (dash-dot line) in Fig. 1.8a, b, respectively.

1.3.5.4 PHYDYAS Prototype Filter

This filter designed by Bellanger in [2, 14] has been considered as the reference prototype filter of the European project PHYDYAS.[2] The design of this prototype filter is based on the so-called frequency sampling technique. The analytical formulas for computing the coefficients of this filter provide a wide choice of the main parameters: the number of subchannels (N), the overlapping factor K and roll-off parameter α. In the design of this filter $\alpha = 1$, which means that the transition bands of a subchannel end at the centers of the adjacent subchannels. This means that only immediately adjacent subchannels are significantly interacting with each other.

The design starts with the determination of $L = KN$ desired values $F(k/L); k = 0, \ldots, L-1$ in the frequency domain given such as,

$$F_0 = 1, F_1 = 0.971960, F_2 = 1/\sqrt{2}$$

$$F_3 = \sqrt{1 - F_1^2}, F_k = 0; 4 < k < L-1 \tag{1.42}$$

Then, the prototype filter coefficients are obtained by IDFT as

$$f(t) = \begin{cases} \frac{1}{\sqrt{A}} \left[1 + 2 \sum_{k=1}^{K-1} (-1)^k F_k \cos\left(\frac{2\pi}{KT} kt\right) \right] & t \in [0, KT] \\ 0 & elsewhere \end{cases} \tag{1.43}$$

where A is the normalization factor

$$A = \int_0^{KT} \left[1 + 2 \sum_{k=1}^{K-1} (-1)^k F_k \cos\left(\frac{2\pi}{KT} kt\right) \right]^2 dt = KT \left[1 + 2 \sum_{k=1}^{K-1} F_k^2 \right] \tag{1.44}$$

The time and frequency responses of PHYDYAS prototype filter are shown (dashed line) in Fig. 1.8a, b, respectively.

---

[2]PHYDYAS: EC funded Physical Layer for Dynamic Spectrum Access and Cognitive Radio, http://www.ict-phydas.org/

1.3.5.5 Other Prototype Filters

The authors in [46] have presented an interesting survey on the design of prototype filters for filter bank based multicarrier systems. The filters are investigated by classifying them under three main categories:

Time-limited filters Their impulse responses are non-zero for finite durations as the name implies. The frequency responses of these filters have infinite length due to the finite durations. Since they are specifically designed for finite duration, they provide easy implementations of the systems. Among others, we find in this category: the rectangular window, Hamming, Hanning and Blackman Filters [21], Optimal Finite Duration Pulses/Prolate [21] and Kaiser Filter [32].

Band-limited filters These filters are theoretically defined with limited bandwidth support. However, infinite impulse responses are theoretically required in time domain. Some examples of the band-limited filters are: raised cosine, root-raised cosine, half cosine and half sine filters.

Localized filters They are simultaneously time-limited and band-limited filters e.g. PHYDYAS. We find also isotropic filters which provide the compactness in both time and frequency domains on an equal footing. They can give equal responses in both time and frequency domains. They tend to behave as a Gaussian pulse because of its maximum compactness e.g. Gaussian, Hermite [34] and IOTA filters.

1.3.5.6 Transmultiplexer Impulse Response

The PHYDYAS and IOTA transmultiplexer (transmitter and receiver connected back to back) impulse responses are respectively given in Table 1.1. Note that the time index n is equivalent to $nT/2$.

Each row of the table represents a sequence of the impulse response corresponding to a given subchannel. Row k_0 is the subchannel response that satisfies the Nyquist criteria. The other rows correspond to the impulse responses of the interference filters due to the overlapping of the neighboring subchannels $\{k_0 - 3, k_0 - 2, k_0 - 1, k_0 + 1, k_0 + 2, k_0 + 3\}$.

One can see that the intrinsic interference terms appearing in the neighborhood of the single transmitted impulse are purely imaginary. Hence using the OQAM technique, we can transmit a real PAM symbol every half period $T/2$ instead of a complex QAM symbol every T.

Moreover from this table, one can see that PHYDYAS-FBMC system offers better frequency localization compared to IOTA. On the other hand, IOTA is well localized in the time domain compared to PHYDYAS. Analyzing Table 1.1, we can also see that the IOTA-FBMC response is isotropic in the time-frequency domain.

Table 1.1 Transmultiplexer response of the FBMC system using PHYDYAS and IOTA prototype filters, respectively

PHYDYAS

	n_0-4	n_0-3	n_0-2	n_0-1	n_0	n_0+1	n_0+2	n_0+3	n_0+4
k_0-3	0	0	0	0	0	0	0	0	0
k_0-2	0	$0.0006j$	$-0.0001j$	0	0	0	$-0.0001j$	$0.0006j$	0
k_0-1	$0.0054j$	$0.0429j$	$-0.1250j$	$-0.2058j$	$0.2393j$	$0.2058j$	$-0.1250j$	$-0.0429j$	$0.0054j$
k_0	0	$-0.0668j$	$0.0002j$	$0.5644j$	**1**	$0.5644j$	$0.0002j$	$-0.0668j$	0
k_0+1	$0.0054j$	$-0.0429j$	$-0.1250j$	$0.2058j$	$0.2393j$	$-0.2058j$	$-0.1250j$	$0.0429j$	$0.0054j$
k_0+2	0	$0.0006j$	$-0.0001j$	0	0	0	$-0.0001j$	$0.0006j$	0
k_0+3	0	0	0	0	0	0	0	0	0

IOTA

	n_0-4	n_0-3	n_0-2	n_0-1	n_0	n_0+1	n_0+2	n_0+3	n_0+4
k_0-3	$-0.0001j$	$0.0004j$	$0.0016j$	$-0.0103j$	$-0.0182j$	$0.0103j$	$0.0016j$	$-0.0004j$	$-0.0001j$
k_0-2	0	$0.0016j$	0	$-0.0381j$	0	$-0.0381j$	0	$0.0016j$	0
k_0-1	$0.0013j$	$-0.0103j$	$-0.0380j$	$0.2280j$	$0.4411j$	$-0.2280j$	$-0.0380j$	$0.0103j$	$0.0013j$
k_0	$-0.0007j$	$-0.0183j$	0	$0.4411j$	**1**	$0.4411j$	0	$-0.0183j$	$-0.0007j$
k_0+1	$0.0013j$	$0.0103j$	$-0.0380j$	$-0.2280j$	$0.4411j$	$0.2280j$	$-0.0380j$	$-0.0103j$	$0.0013j$
k_0+2	0	$0.0016j$	0	$-0.0381j$	0	$-0.0381j$	0	$0.0016j$	0
k_0+3	$-0.0001j$	$-0.0004j$	$0.0016j$	$0.0103j$	$-0.0182j$	$-0.0103j$	$0.0016j$	$0.0004j$	$-0.0001j$

1.4   Spectrum Sharing and Coexistence: FBMC Application

Thanks to their numerous advantages, multicarrier techniques are considered as potential candidates for the physical layer technology for dynamic spectrum access and cognitive radio systems [27]. Also, Fourrier analysis used in the OFDM receiver is suitable for spectrum sensing, which is one of the principal functions in the cognitive radio concept. It allows in fact the determination of the white (unused) frequency bands [62].

Despite theses favorable characteristics, conventional OFDM exhibits serious limitations in cognitive radio contexts. One of these main issues is the poor spectral containment of the CP-OFDM subcarriers, leading to high sidelobes in the modulated OFDM spectrum, which creates interference to adjacent systems frequencies and is thus problematic for the coexistence scenario. Such a problem prevents efficient utilization of the free frequency bands by cognitive users which strongly limits the overall spectral efficiency of cognitive radio networks. Furthermore, it has been demonstrated that OFDM performances are tremendously sensitive to timing synchronization errors [42] which inherently exist in cognitive system.

These drawbacks can be avoided with filtered OFDM. However, this latter has some serious weak points. In this case, the subcarrier synthesis and analysis filters, respectively, suffer from significantly wide transition bands [21] due to the small roll-off periods of these filters.

As previously mentioned, cognitive networks are assumed to be flexible where primary and secondary users independently transmit and may be based on different standards. In this case, the utilization of FBMC systems becomes more advantageous and convenient due to their filtering properties [21]. In other words, the good spectral localization of FBMC waveforms make them more suitable for cognitive radio networks physical layer allowing an efficient opportunistic spectrum access. Besides, a reduced implementation complexity can be achieved by relaxing all synchronization mechanisms between the primary and the secondary systems thanks to the robustness of FBMC waveforms to synchronization errors [41].

In order to demonstrate the detrimental effects of interference caused by the imperfect synchronization in a coexistence spectrum scenario, let us consider the system model depicted in Fig. 1.9. The reference mobile user MU_0 and the interfering one MU_1 communicating respectively with BS_0 and BS_1. Moreover, MU_0 and MU_1 are respectively located at distances d_0 and d from the reference base station BS_0. It is assumed that, the transmitted power of each user must guarantee a target signal to noise ration $SNR_t = 20\,dB$ at its base station (BS_0 for MU_0 and BS_1 for MU_1).

Concerning the frequency scheme, the subcarriers are allocated according to the scheme described in Fig. 1.9b, where, the size of each subcarrier block is set at 18 subcarriers. Here, we have chosen the practical size of subcarrier block in WiMax 802.16 [15].

All signals propagate through different multipath channels using a similar propagation model, where the impulse responses of the multipath channel between MU_0,
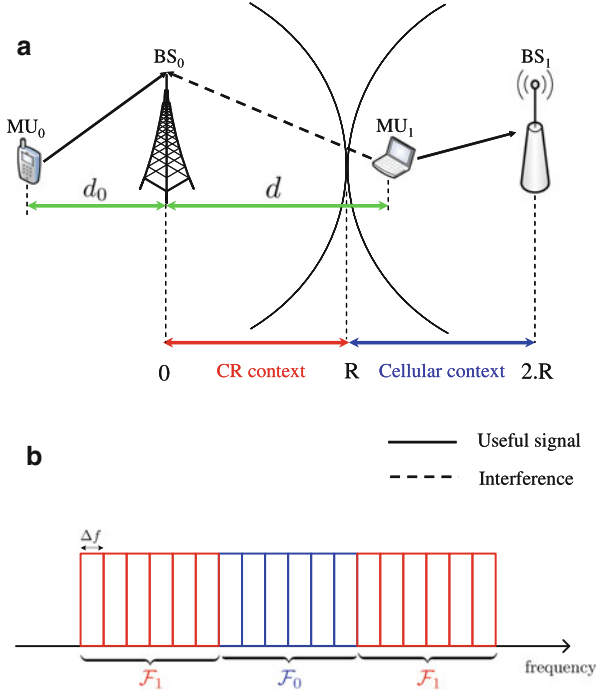
Fig. 1.9 Interference model: the reference user coexists with an asynchronous interferer

Table 1.2 Channel
parameters used in
simulations

Parameter	Value
Pedestrian-A relative delay	[0 110 190 410] ns
Pedestrian-A average power	[0 −9.7 −19.2 −22.8] dB

MU_1 and the reference base station BS_0 are denoted by h_0 and h_1, respectively. The considered model is the Pedestrian-A model whose parameters are given in Table 1.2. The choice of this model is based on the assumption that the subcarriers of interest experience flat fading. Therefore, the interference caused by the multipath effects are negligible, in the FBMC case.

Furthermore, the underlying channel model includes path loss effects which takes into account the position of the mobile user with respect to the reference base station BS_0. The path loss of a received signal at distance d is governed by the following expression [18] corresponding to a path loss exponent $\beta = 3.76$ and a carrier frequency of 2 GHz

$$\Gamma_{\text{loss}}(d) = 128.1 + 37.6 \log_{10}(d \, [\text{km}]) [\text{dB}]$$

On the other hand, we consider a system with $N = 1,024$ subcarriers and a sampling frequency of 10 MHz. The noise term is considered as a thermal noise with spectral density $N_0 = -174$ dBm/Hz.
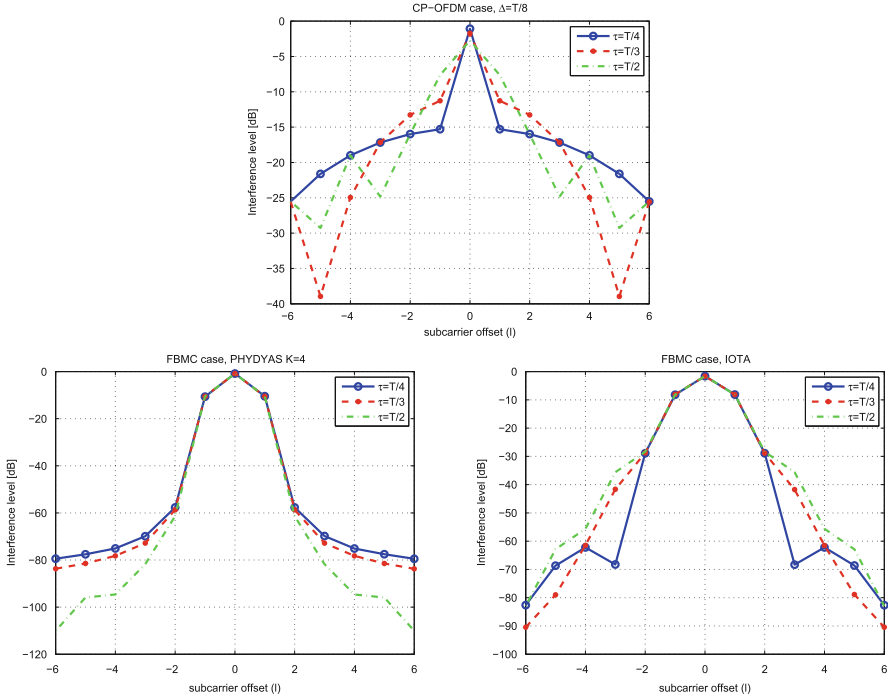
Fig. 1.10 CP-OFDM, PHYDYAS and IOTA instantaneous interference tables for $\tau = \{T/4, T/3, T/2\}$

Each mobile user is assumed to be perfectly synchronized with its corresponding base station but it is not synchronized with the other base station. Because of the timing misalignment between MU_1 and BS_0, the signal arriving from MU_1 at BS_0 appears non-orthogonal to the desired signal arriving from MU_0. This non-orthogonality generates interference and degrades the signal to interference plus noise ratio SINR. In order to evaluate this SINR degradation, a useful tool called "instantaneous interference tables" has been introduced in [39]. These tables are able to give the interference power $I(\tau, |m' - m|)$ caused by a given interfering subcarrier m' on a victim one m as a function of two parameters: the spectral distance separating both subcarriers $|m'-m|$ and the timing misalignment τ between the subcarrier holders.

In Fig. 1.10, CP-OFDM, PHYDYAS and IOTA interference tables $I(\tau, l)$ are plotted for different values of the timing offset $\tau = T/4, T/3, T/2$. One can see that the interference varies with respect to the timing offset τ and the spectral distance l between the interfering subcarrier and the victim one.

Therefore using these tables, we can express the instantaneous SINR on a given subcarrier $m \in \mathcal{F}_0$ in the following form,

$$\text{SINR}(m) = \frac{d_0^{-\beta} P_{\text{trans}}(m) \, |H_0(m)|^2}{\sum\limits_{m' \in \mathcal{F}_1} d^{-\beta} P_{\text{trans}}(m') I(\tau, |m' - m|) \, |H_1(m')|^2 + N_0 \Delta f} \qquad (1.45)$$

where Δf is the subcarrier spacing.

Now, our aim is to evaluate the average SINR and the spectral efficiency expressed, respectively by,

$$\text{SINR}_{\text{average}}(m) = \mathbb{E}\left[\text{SINR}(m)\right] \qquad (1.46)$$

$$C_{\text{average}}(m) = \mathbb{E}\left[\log_2\left(1 + \text{SINR}(m)\right)\right] \qquad (1.47)$$

where $\mathbb{E}[.]$ stands for the statistical expectation which is computed over all channel realizations $(H_0(m), \{H_1(m'), m' \in \mathcal{F}_1\})$ and all values of the timing offset τ which is uniformly distributed over $[0, T]$.

Two different contexts will be analyzed as depicted in Fig. 1.9:

- **The classical multi-cellular context**: when d varies from R to $2R$, i.e. MU_1 can move from the edge to the center of cell 1
- **The cognitive radio context**: when d varies from 0 to $2R$, i.e. MU_1 can be very close to BS_0 while transmitting to BS_1.

The OFDM, PHYDYAS-FBMC and IOTA-FBMC SINRs averaged over all subcarriers $m \in \mathcal{F}_0$ are plotted against the distance d, in Fig. 1.11.

In the cognitive radio context, we observe a significant degradation of the OFDM SINR with respect to the target SNR (20 dB). Such a result can be explained by the high level of OFDM asynchronous interference caused by the timing misalignment which damages the orthogonality between the subcarriers. On the other hand, we notice a slight loss of the FBMC SINR with respect also to the target SNR of 20 dB. The better performance of PHYDYAS-FBMC compared to IOTA-FBMC and CP-OFDM can be justified by the fact that only the two subcarriers on the edge of the cluster (subcarrier block) $\mathcal{F}_0$ suffer from the interference caused by their immediate adjacent subcarriers in $\mathcal{F}_1$ as depicted in Fig. 1.10; whereas in IOTA-FBMC, two subcarriers at each edge are affected by the asynchronous interference coming from the two neighboring subcarriers at each edge as shown in Fig. 1.10. Furthermore, the entire cluster $\mathcal{F}_0$ suffers, in the CP-OFDM case, from the asynchronous interference caused by all subcarriers of $\mathcal{F}_1$ (see Fig. 1.10).

In the cellular context, the asynchronous interferer MU_1 is quite far from the reference base station BS_0 and, at the same time, it is close to its base station BS_1. This means that its transmitted power is reduced and consequently the interference power received by BS_0 will be much lower due to the path-loss effect (d is quite large). Therefore, the impact of the asynchronous interference is less significant in the cellular context for all waveforms. Also, it is worth mentioning that SINRs of OFDM and FBMC converge to the target SNR (20 dB) when MU_1 is very far from BS_0 as the interference becomes negligible compared to the noise level.

The impact of the asynchronous interference on the average spectral efficiency has also been investigated. Figure 1.12, shows the average spectral efficiency over all
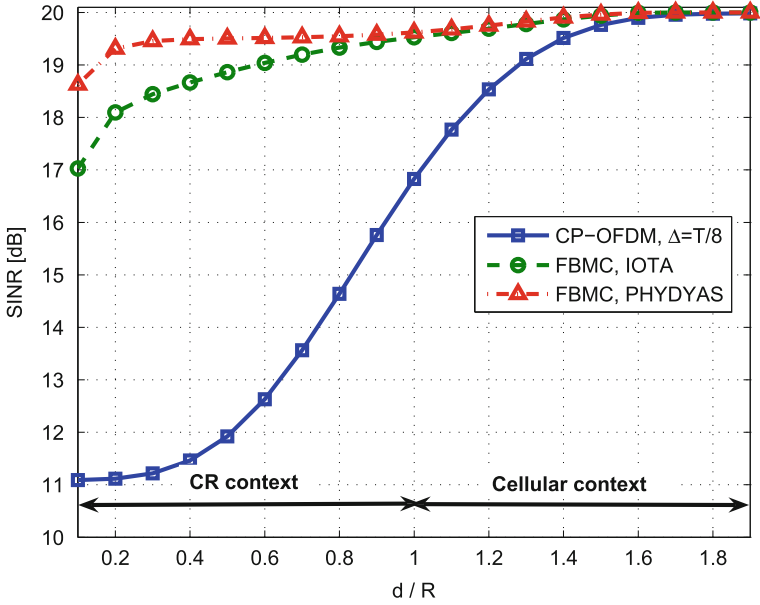
Fig. 1.11 CP-OFDM, IOTA-FBMC and PHYDYAS-FBMC average SINR vs. distance between MU_1 and BS_0, $\tau \in [0, T]$
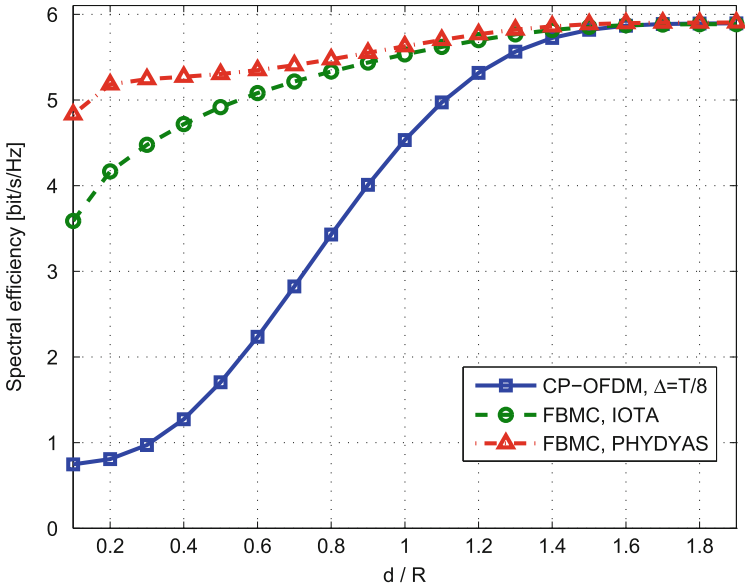
Fig. 1.12 CP-OFDM, IOTA-FBMC and PHYDYAS-FBMC spectral efficiency vs. distance between MU_1 and BS_0, $\tau \in [0, T]$

subcarriers $m \in \mathcal{F}_0$ against the distance d, for CP-OFDM, FBMC-PHYDYAS and FBMC-IOTA. The obtained results show that timing synchronization errors cause a degradation of the spectral efficiency. The same remarks can be formulated, here, where the FBMC system still outperforms the OFDM system. We see that both modulation schemes (OFDM and FBMC) lead to identical spectral efficiency floor when MU_1 is close to its base station because of the predominance of the noise term. We have to remind that the actual bit rate is lower for CP-OFDM because of the redundancy introduced by the cyclic prefix in each OFDM block.

1.5   Conclusion

In this chapter, we have presented the main framework of filter bank based multicarrier techniques. We have first given a brief review of the general concept of multicarrier transmission. We then discussed the advantages and the shortcomings of the conventional OFDM technique. The second part of this chapter was devoted to describe FBMC techniques. The different forms of FBMC: FMT, CMT and SMT (FBMC/OQAM) were introduced. Furthermore, we have shown that FBMC introduces extra complexity by adding well-localized filters at the transmitter and the receiver. However, this complexity can be reduced by using the efficient polyphase implementation of the synthesis filter bank (FBMC transmitter) and the analysis filter bank (FBMC receiver). The choice and the design of the prototype filters that can be used in FBMC have also been discussed. In the last part of this chapter, we have presented an overview on the interest of FBMC in spectrum sharing and coexistence issues in cognitive radio networks. The favorable properties of FBMC such as: its good spectral containment and its robustness to synchronization errors have been highlighted showing then the suitability of FBMC for cognitive radio applications.

References

1. Alard, M: Construction of a multicarrier signal (Patent WO/1996/035278) (1996). Available in WIPO: http://patentscope.wipo.int/search/en/WO1996035278. Accessed Jun 2013
2. Bellanger, M.: Specification and design of a prototype filter for filter bank based multicarrier transmission. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01), Salt Lake City, vol. 4, pp. 2417–2420 (2001)
3. Bellanger, M.: FBMC physical layer: a primer. Technical report, European project ICT-211887 PHYDYAS (2010)
4. Bellanger, M., Bonnerot, G., Coudreuse, M.: Digital filtering by polyphase network: application to sample-rate alteration and filter banks. IEEE Trans. Acoust. Speech Signal Process. **24**(2), 109–114 (1976)

5. Bellanger, M., Daguet, J.: TDM-FDM transmultiplexer: digital polyphase and FFT. IEEE Trans. Commun. **22**(9), 1199–1205 (1974)
6. Benvenuto, N., Cherubini, G., Tomba, L.: Achievable bit rates of DMT and FMT systems in the presence of phase noise and multipath. In: 2000 IEEE 51st Vehicular Technology Conference Proceedings (VTC 2000-Spring), Tokyo, vol. 3, pp. 2108–2112 (2000). doi:10.1109/VETECS.2000.851644
7. Bölcskei, H.: Orthogonal frequency division multiplexing based on offset QAM. In: Advances in Gabor Analysis, pp. 321–352. Birkhäuser, Boston (2003)
8. Cariolaro, G., Vagliani, F.: An OFDM scheme with a half complexity. IEEE J. Sel. Areas Commun. **13**(9), 1586–1599 (1995)
9. Chang, R.W.: High-speed multichannel data transmission with bandlimited orthogonal signals. Bell Syst. Tech. J. **15**(6), 1775–1796 (1966)
10. Cherubini, G., Eleftheriou, E., Oker, S., Cioffi, J.: Filter bank modulation techniques for very high speed digital subscriber lines. IEEE Commun. Mag. **38**(5), 98–104 (2000). doi:10.1109/35.841832
11. Cherubini, G., Eleftheriou, E., ölçer, S.: Filtered multitone modulation for very high-speed digital subscriber lines. IEEE J. Sel. Areas Commun. **20**(5), 1016–1028 (2002). doi:10.1109/JSAC.2002.1007382
12. Chevillat, P., Ungerboeck, G.: Optimum fir transmitter and receiver filters for data transmission over band-limited channels. IEEE Trans. Commun. **30**(8), 1909–1915 (1982)
13. D2.1, D.: Data-aided synchronization and initialization (single antenna). Technical report, European project ICT-211887 PHYDYAS (2009)
14. D5.1, D.: Prototype filter and structure optimization. Technical report, European project ICT-211887 PHYDYAS (2009)
15. Etemad, K., Lai, M.-Y.: WiMAX Technology and Network Evolution (Piscataway: Wiley, IEEE Press, 2011)
16. ETSI: Radio broadcasting systems: digital audio broadcasting (DAB) to mobile, portable and fixed receivers. ETS 300 401, European Telecommunications Standard Institute (1997)
17. ETSI: Digital video broadcasting (DVB-T); frame structure, channel coding, modulation for digital terrestrial television. ETS 300 744, European Telecommunications Standard Institute (2001)
18. ETSI: Universal mobile telecommunications system (UMTS); radio frequency (RF) system scenarios. ETS TR 125 942, European Telecommunications Standard Institute (2010)
19. Farhang-Boroujeny, B.: Multicarrier modulation with blind detection capability using cosine modulated filter banks. IEEE Trans. Commun. **51**(12), 2057–2070 (2003)
20. Farhang-Boroujeny, B.: Signal Processing Techniques for Software Radios. LuLu Publishing, Morrisville (2005)
21. Farhang-Boroujeny, B.: OFDM versus filter bank multicarrier. IEEE Signal Process. Mag. **28**(3), 92–112 (2011)
22. Feichtinger, H., Strohmeri, T.: Gabor Analysis and Algorithms: Theory and Applications. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston (1998)
23. Fliege, N.J.: Multirate Digital Signal Processing: Multirate Systems, Filter Banks, Wavelets. Wiley, Hoboken (1994)
24. Goldsmith, A.: Wireless Communications. Cambridge University Press, New York (2005)
25. Group, I..W.: Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: higher-speed physical layer extension in the 5 GHz band. IEEE802.11a (1999)
26. Hamdi, K., Shobowale, Y.: Interference analysis in downlink OFDM considering imperfect intercell synchronization. IEEE Trans. Veh. Technol. **58**(7), 3283–3291 (2009)
27. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005). doi:10.1109/JSAC.2004.839380
28. Hirosaki, B.: An analysis of automatic equalizers for orthogonally multiplexed QAM systems. IEEE Trans. Commun. **28**(1), 73–83 (1980)
29. Hirosaki, B.: An orthogonally multiplexed QAM system using the discrete Fourier transform. IEEE Trans. Commun. **29**(7), 982–989 (1981)

30. Hoeg, W., Lauterbach, T.: Digital Audio Broadcasting: Principles and Applications of DAB, DAB+ and DMB. Wiley, Chichester (2009)
31. Ihalainen, T., Stitz, T.H., Rinne, M., Renfors, M.: Channel equalization in filter bank based multicarrier modulation for wireless communications. EURASIP J. Adv. Signal Process. **2007**(1), 18 (2007)
32. Kaiser, J., Schafer, R.: On the use of the i0-sinh window for spectrum analysis. IEEE Trans. Acoust. Speech Signal Process. **28**(1), 105–107 (1980). doi:10.1109/TASSP.1980.1163349
33. Kozek, W., Molisch, A.: Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels. IEEE J. Sel. Areas Commun. **16**(8), 1579–1589 (1998). doi:10.1109/49.730463
34. Kurt, T., Siala, M., Yongacoglu, A.: Multi-carrier signal shaping employing Hermite functions. In: Proceedings of the European Signal Processing Conference (EUSIPCO), Antalya (2005)
35. Le Floch, B., Alard, M., Berrou, C.: Coded orthogonal frequency division multiplex [TV broadcasting]. Proc. IEEE **83**(6), 982–996 (1995)
36. Lélé, C.: OFDM/OQAM: méthodes d'estimation de canal, et combinaison avec l'accès multiple CDMA ou les systèmes multi-antennes. Ph.D. thesis, Conservatoire National des Arts et Métiers, Paris (2008)
37. Lin, H., Siohan, P., Tanguy, P., Javaudin, J.P.: An analysis of the EIC method for OFDM/OQAM systems. J. Commun. **4**(1), 52–60 (2009)
38. Martin, K.: Small side-lobe filter design for multitone data-communication applications. IEEE Trans. Circuits Syst. II Analog Digit. Signal Process. **45**(8), 1155–1161 (1998). doi:10.1109/82.718830
39. Medjahdi, Y., Terré, M., Le Ruyet, D., Roviras, D.: Interference tables: a useful model for interference analysis in asynchronous multi-carrier transmission. EURASIP J. Adv. Signal Process. (1), 17 (2014)
40. Medjahdi, Y., Terre, M., Le Ruyet, D., Roviras, D.: Asynchronous OFDM/FBMC interference analysis in selective channels. In: 2010 IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Istanbul, pp. 538–542 (2010). doi:10.1109/PIMRC.2010.5671895
41. Medjahdi, Y., Terre, M., Le Ruyet, D., Roviras, D.: On spectral efficiency of asynchronous OFDM/FBMC based cellular networks. In: 2011 IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Toronto, pp. 1381–1385 (2011). doi:10.1109/PIMRC.2011.6139729
42. Medjahdi, Y., Terre, M., Le Ruyet, D., Roviras, D., Dziri, A.: Performance analysis in the downlink of asynchronous OFDM/FBMC based multi-cellular networks. IEEE Trans. Wirel. Commun. **10**(8), 2630–2639 (2011). doi:10.1109/TWC.2011.061311.101112
43. Morelli, M., Kuo, C.C., Pun, M.O.: Synchronization techniques for orthogonal frequency division multiple access (OFDMA): a tutorial review. Proc. IEEE **95**(7), 1394–1427 (2007)
44. Nedic, S.: An unified approach to equalization and echo cancellation in OQAM-based multi-carrier data transmission. In: IEEE Global Telecommunications Conference, Phoenix, vol. 3, pp. 1519–1523 (1997)
45. Proakis, J.: Digital Communications, 4th edn. McGraw-Hill, New York (1994)
46. Sahin, A., Güvenç, I., Arslan, H.: A survey on prototype filter design for filter bank based multicarrier communications (2012). arXiv [cs.IT] abs/1212.3374
47. Saltzberg, B.: Performance of an efficient parallel data transmission system. IEEE Trans. Commun. Technol. **15**(6), 805–811 (1967)
48. Sandberg, S., Tzannes, M.: Overlapped discrete multitone modulation for high speed copper wire communications. IEEE J. Sel. Areas Commun. **13**(9), 1571–1585 (1995)
49. Siohan, P., Roche, C.: Cosine-modulated filterbanks based on extended Gaussian functions. IEEE Trans. Signal Process. **48**(11), 3052–3061 (2000)
50. Siohan, P., Siclet, C., Lacaille, N.: Analysis and design of OFDM/OQAM systems based on filter bank theory. IEEE Trans. Signal Process. **50**(5), 1170–1183 (2002)
51. Sondergaard, P.L.: Finite discrete Gabor analysis. Ph.D. thesis, Danmarks Tekniske Universitety (2007)

52. Strohmer, T., Beaver, S.: Optimal OFDM design for time-frequency dispersive channels. IEEE Trans. Commun. **51**(7), 1111–1122 (2003)
53. Tse, D., Viswanath, P.: Fundamentals of Wireless Communications. Cambridge University Press, New York (2004)
54. Vahlin, A., Holte, N.: Optimal finite duration pulses for OFDM. IEEE Trans. Commun. **44**(1), 10–14 (1996)
55. Vaidyanathan, P.P.: Multirate Systems and Filter Banks. Prentice-Hall, Upper Saddle River (1993)
56. Van Acker, K., Leus, G., Moonen, M., van de Wiel, O., Pollet, T.: Per Tone equalization for DMT-based systems. IEEE Trans. Commun. **49**(1), 109–119 (2001)
57. Vandendorpe, L., Cuvelier, L., Deryck, F., Louveaux, J., van de Wiel, O.: Fractionally spaced linear and decision-feedback detectors for transmultiplexers. IEEE Trans. Signal Process. **46**(4), 996–1011 (1998)
58. Vangelista, L., Laurenti, N.: Efficient implementations and alternative architectures for OFDM-OQAM systems. IEEE Trans. Commun. **49**(4), 664–675 (2001)
59. Waldhauser, D., Baltar, L., Nossek, J.: MMSE subcarrier equalization for filter bank based multicarrier systems. In: IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, Recife, pp. 525–529 (2008)
60. Waldhauser, D., Baltar, L., Nossek, J.: Adaptive decision feedback equalization for filter bank based multicarrier systems. In: IEEE International Symposium on Circuits and Systems, Taipei, pp. 2794–2797 (2009)
61. Weiss, T., Hillenbrand, J., Krohn, A., Jondral, F.: Mutual interference in OFDM-based spectrum pooling systems. In: IEEE 59th Vehicular Technology Conference, Milan, vol. 4, pp. 1873–1877 (2004)
62. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009). doi:10.1109/SURV.2009.090109

Chapter 2
Cognitive Interference Alignment for Spectral Coexistence

Shree Krishna Sharma, Symeon Chatzinotas, and Björn Ottersten

Abstract Interference Alignment (IA) has been widely recognized as a promising interference mitigation technique since it can achieve the optimal degrees of freedom in certain interference limited channels. In the context of Cognitive Radio (CR) networks, this technique allows the coexistence of two heterogeneous wireless systems in an underlay cognitive mode. The main concept behind this technique is the alignment of the interference on a signal subspace in such a way that it can be filtered out at the non-intended receiver by sacrificing some signal dimensions. This chapter starts with an overview of IA principle, Degree of Freedom (DoF) concept, and the classification of existing IA techniques. Furthermore, this chapter includes a discussion about IA applications in CR networks. Moreover, a generic system model is presented for allowing the coexistence of two heterogeneous networks using IA approach while relevant precoding and filtering processes are described. In addition, two important practical applications of the IA technique are presented along with the numerical results for underlay spectral coexistence of (i) femtocell-macrocell systems, and (ii) monobeam-multibeam satellite systems. More specifically, an uplink IA scheme is investigated in order to mitigate the interference of femtocell User Terminals (UTs) towards the macrocell Base Station (BS) in the spatial domain and the interference of multibeam satellite terminals towards the monobeam satellite in the frequency domain.

S.K. Sharma (✉) • S. Chatzinotas • B. Ottersten
SnT, University of Luxembourg, L-2721, Kirchberg, Luxembourg
e-mail: shree.sharma@uni.lu; symeon.chatzinotas@uni.lu; bjorn.ottersten@uni.lu

2.1   Introduction

The demand for the broadband wireless spectrum is increasing due to rapidly increasing number of broadband and multimedia wireless users and applications. Due to the limited and expensive frequency resources, Cognitive Radio (CR) communication can be an efficient technique to enhance the spectrum efficiency since it allows the coexistence of primary and secondary wireless networks within the same spectrum. Wireless networks may coexist within the same spectrum band in different ways such as two terrestrial networks, two satellite networks or satellite-terrestrial networks. Due to the recent advancements in terrestrial cellular technology and multibeam satellite technology, denser deployments of cells/beams have become possible for providing higher capacity and network availability. In the context of terrestrial systems, small cell systems such as femtocells have received important attention due to higher cellular capacity and energy efficiency harnessed by switching unused femtocells in a sleep mode [3]. Furthermore, femtocells can provide better user experience with lower capital and operational costs compared to other techniques for indoor coverage. Similarly, in the satellite paradigm, multiple beams can be employed instead of a single global beam in order to enhance the capacity [12]. However, current network configurations use large cell systems and the deployment of new small cell systems need additional bandwidth which is scarce and expensive to acquire. In this context, dense small cell systems have to coexist with the traditional large cell systems to optimally utilize the existing spectrum.

Interference is an inevitable phenomenon in wireless communication systems when multiple uncoordinated links share a common wireless channel. The coexistence of different wireless networks in the same spectrum band can be modeled as CR networks with interference channels between primary and secondary systems. The operation of the primary network usually follows a well established standard and should not be degraded while the secondary network should employ some advanced transmission and coding techniques in order to exploit the underutilized dimensions in the frequency, time and space domains. Depending on the strength of the interference between wireless networks, different interference management approaches can be applied. If the interference is weaker than the noise floor, the interference signal can be treated as noise and the single user encoding/decoding mechanisms can be applied. Because of its simplicity and ease of implementation, this approach is widely used in practice, but does not achieve interference-free capacity even for the simple case of a Broadcast Channel (BC) [71]. If the interference level is strong in comparison to the noise floor, it is possible to decode the interference and then subtract it from the received signal. This method is less common in practice due to its complexity and security issues. However, when the strength of the interference is comparable to the desired signal, treating as noise is not an option because of interference constraints involved while decoding and canceling requires complex primary receivers. In this case, one approach is to orthogonalize channels so that transmitted signals are chosen to be non-overlapping in the time, frequency or space domain, leading to Time Division Multiple Access

(TDMA), Frequency Division Multiple Access (FDMA) or Space Division Multiple Access (SDMA) respectively. Furthermore, in multiuser interference networks, applying the above techniques is problematic since the aggregate interference may be stronger than the noise floor in many cases and decoding may also be complex due to involvement of several interfering users. Although the orthogonalization approach effectively eliminates multiuser interference in wireless networks, it may lead to underutilization of communication resources and it also does not achieve the capacity of interference channels [51]. In this context, Interference Alignment (IA) has received important attention as an interference mitigation tool in interference-limited wireless systems such as cellular wireless networks, CR systems and ad-hoc networks.

The remainder of this chapter is structured as follows: Sect. 2.2 introduces the fundamentals of the IA technique including Degrees of freedom (DoF) concept, basic IA principle and the classification of IA techniques. Section 2.3 includes the current state of art related to the application of IA in CR networks. Section 2.4 includes the generic system model for spectrum coexistence scenario in which IA technique can be applied and further describes the mechanism for IA and filtering process. Section 2.5 provides the application of different IA approaches for the following two practical scenarios including numerical results: (i) femtocell-macrocell coexistence scenario, and (ii) monobeam-multibeam satellite coexistence. Section 2.6 presents the challenges of IA technique from practical perspectives and further includes future research directions. Section 2.7 summarizes the chapter.

2.1.1 Notation

Throughout this chapter, boldface upper and lower case letters are used to denote matrices and vectors respectively, $\mathbb{E}[\cdot]$ denotes expectation, $(\cdot)^{\dagger}$ denotes the conjugate transpose matrix, $(\cdot)^{T}$ denotes the transpose matrix, $O(\cdot)$ denotes the order, $(z)^{+}$ denotes max$(0, z)$, and $\mathbf{0}$ represents a zero matrix

2.2 Interference Alignment (IA) Fundamentals

In wireless interference networks, only a subset of the transmitted symbols are desired by a particular receiver. The remaining symbols, which carry information for other receivers, are undesired at that particular receiver creating interference to the desired signal. In this context, IA can be used as an interference mitigation tool which aligns interference in the space, time or frequency domain using precoding techniques. The main principle behind IA is the alignment of the interference on a signal subspace in such a way that it can be easily filtered out at the non-intended receiver by sacrificing some signal dimensions. In other words, signals transmitted by all users can be designed in such a way that the interfering signals

fall into a reduced dimensional subspace at each receiver. Each receiver can then apply an interference removal filter to project the desired signal onto the interference free subspace. Due to this approach, the number of interference-free signalling dimensions of the network are substantially increased [27]. In Multiple Input Multiple Output (MIMO) networks, IA can be applied by using the spatial dimension offered by multiple antennas for alignment while in multicarrier systems, interference can be aligned along the carrier dimension.

2.2.1 Degrees of Freedom (DoF)

The DoF is an important metric used for capacity approximation in wireless networks literature. It may be interpreted as the number of resolvable signal space dimensions and is a way of measuring the spatial multiplexing gain provided by MIMO systems at high Signal to Noise Ratios (SNRs). It can also be defined as the number of signaling dimensions, each dimension corresponding to one interference-free Additive White Gaussian Noise (AWGN) channel with Signal to Noise Ratio (SNR) that increases proportionally with the total transmit power P as $P \rightarrow \infty$ [27]. The DoF also corresponds to the multiplexing gain, bandwidth, capacity pre-log factor, or the number of signaling dimensions. Let $R(P)$ denotes the sum capacity, then the DoF metric, let us denote by η, is given by

$$\eta = \lim_{P \to \infty} \frac{R(P)}{\log(P)}. \tag{2.1}$$

The above expression can be equivalently written as: $R(P) = \eta \log(P) + O(\log(P))$, where the term $O(\log(P))$ is some function $f(P)$ which satisfies the following relation [27]

$$\lim_{P \to \infty} \frac{f(P)}{\log(P)} = 0. \tag{2.2}$$

For example, a point to point MIMO channel with M transmit and N receive antennas has $\min(M, N)$ DoF, whereas it's Single Input Single Output (SISO) counterpart has only 1 DoF [65].

The DoF regions are characterized for several wireless channels such as MIMO BC, interference channels (ICs), including X and multihop ICs, and the CR channels [68]. The DoF metric has been extensively used for interference mitigation and alignment objectives in various wireless networks such as interference mitigation in multicell networks [10, 31], interference mitigation in two-cell MIMO interfering BCs [63], IA in CR networks [14, 34, 57]. The main limitation of the DoF metric is that it does not provide much insight to optimally manage interference when all signals are not comparable, since it forces all channels to be equally strong. In this case, another metric, called Generalized DoF (GDoF), can be used [4]. This metric

can preserve the diversity of signal strengths by fixing the ratios of different signal powers when all SNRs approach infinity. Let α denote the ratio of the cross channel strength to the direct channel strength in dB scale and $R(P, \alpha)$ denote the sum-capacity. Then the total GDoF metric, $\eta(\alpha)$ can be defined as [27]:

$$\eta(\alpha) = \lim_{P \to \infty} \frac{R(P, \alpha)}{\log(P)}. \tag{2.3}$$

It can be noted that the GDoF metric corresponds to the DoF metric when $\alpha = 1$. This metric has been successfully used in [4, 18] to approximate the capacity of two user interference channel and in [46] for multiple antenna scenarios considering two user MIMO interference channel.

2.2.2 IA Principle

The IA technique allows many interfering users to communicate simultaneously over a small number of signaling dimensions i.e., number of antennas or carriers. This is achieved by aligning the space spanned by the interference at each receiver within a small number of dimensions and keeping the desired signals distinguishable from interference so that they can be projected into null space of the interference and desired signal can be recovered from the received signal. The disadvantage of the IA approach is that filtering at the non-intended receiver removes the signal energy in the interference subspace. Let us consider an interference network with K transmitters, each trying to send one information symbol. To resolve the 1 symbol desired by a particular receiver, K signalling dimensions are generally required [27]. If there are K number of receivers, each with access to a different set of K linear equations formed by its linear channel to the transmitters and interested in a different symbol, a total number of K signalling dimensions will be sufficient to recover the desired symbol by all the K receivers. In this case, the total signalling dimensions are shared among the K users so that each user can communicate using $1/K$ fraction of it like a cake-cutting bandwidth allocation. If all the available receiving dimensions are spanned by interference beams, the desired signal will lie within the interference space as well and can not be resolved. However, if the signals can be designed in such a way that the interference beams can be consolidated into a smaller subspace i.e., they do not span the entire available signal space at the receiver, and the desired signal beam can avoid falling into the interference space, then the receiver becomes able to recover its desired symbol. The advantage of this mitigation approach is that this alignment does not affect the randomness of the signals and the available dimensions with respect to the intended receiver. The fundamental assumptions which make IA feasible are that there are multiple available dimensions (space, frequency, time or code) and that the transmitter is aware of the Channel State Information (CSI) towards the non-intended receiver.

The relativity of alignment is an important aspect for enabling IA in interference wireless networks [5]. It implies that when there are multiple non-intended receivers, the alignment of signals in these receivers is different i.e., the set of input-output equations observed in each receiver is different from those observed in other receivers. Since the signals do not align into the desirable patterns naturally, the most important challenge for IA techniques is the design of signal vectors to fulfill the desired alignment conditions, which are explained later in Sect. 2.3 for different IA techniques. In the context of multiple non-intended receivers, applying IA is not straightforward since the alignment for one receiver in general does not ensure alignment at other receivers as well.

For the Gaussian interference channel with K interfering transmitter-receiver pairs with each transmit and receive node having M antennas each, and with random, time-varying channel coefficients drawn from a continuous distribution, the sum-capacity of the network is characterized as [5]:

$$R(P) = \frac{KM}{2}\log(P) + O(\log(P)). \tag{2.4}$$

In this case, capacity per transmit-receive pair, i.e., for one user, becomes $\frac{M}{2}\log(P) + O(\log(P))$, where P is the total transmit power of all the transmitters in the network when the noise power is normalized to unity. The term $O(\log(P))$ becomes negligible as compared to $\log(P)$ at high SNRs and the accuracy of the capacity approximation approaches 100 %. Based on the results obtained in [5], it can be deduced that every user in a wireless interference network is (simultaneously and almost surely) able to achieve approximately one half of the interference-free capacity. From the sum-rate perspective, with K user pairs, an IA strategy achieves the sum throughput on the order of $K/2$ interference-free channels. More specifically, each user can effectively get half the system capacity. Thus in contrast to conventional interference channels, there is increase in the sum rate with the number of active user pairs. To illustrate the IA principle, Fig. 2.1 presents the spectral coexistence scenario of a primary and a secondary cellular networks. The secondary transmitters apply precoding using a predefined or coordinated alignment vector before transmitting so that the interfering signals are all aligned at the primary receiver at a certain direction. Then the received signal at the primary receiver is filtered out by using suitably designed filter so that the interference is filtered out, only leaving the desired signal at the output. The detailed description on this alignment and filtering process is presented in Sect. 2.4.

The main drawback of the IA technique from practical point of view is that it requires the global or local CSI knowledge depending on the applied techniques. The CSI for IA operation can be obtained basically by the following two methods [17].

1. **CSI through Reciprocity**: In Time Division Duplex (TDD) based systems, propagation in both directions can be considered to be identical and the channels are said to be reciprocal. Reciprocity enables the IA by allowing transmitters to
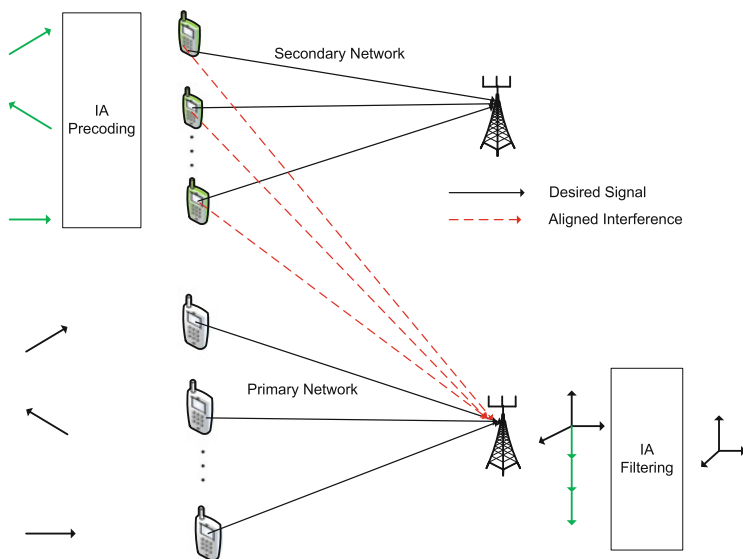
Fig. 2.1 Illustration of IA principle in a cellular network

predict the strength of the interference they cause by observing the interference they receive. The general framework for reciprocity consists of forward link training and reverse link training until the convergence occurs and then the data transmission phase gets started.

2. **CSI through Feedback**: In this approach, a transmitter first sends a training sequence and based on this training sequence, the receiver estimates the forward channel. Subsequently, the receiver feeds back this estimated channel information, potentially after training the reverse link. After feedback, the transmitter has the information needed to design an IA precoder. The disadvantage of this method is that feedback process introduces distortion to the CSI at the transmitter and may create a non-negligible overhead penalty.

2.2.3   Classification of IA Techniques

The IA technique was firstly proposed in [6] and channel capacity as well as DoFs for the interference channel have been analyzed. This technique has been shown to achieve the DoFs for a range of interference channels [5, 7, 28]. Finding out the exact number of needed dimensions and the precoding vectors to achieve IA is a cumbersome task but a number of approaches have been presented in the literature for this purpose [21, 66, 75]. The IA technique was also investigated in the context of cellular networks, showing that it can effectively suppress cochannel interference [9, 15, 64, 66]. More specifically, the downlink of orthogonal frequency division

multiple access (OFDMA) cellular network with clustered multicell processing is considered in [15], where IA is employed to suppress intracluster interference while intercluster interference has to be tolerated as noise. In addition, authors in [64] consider the uplink of a limited-size cellular system without Multicell Joint Decoding (MJD), showing that the interference-free DoFs can be achieved as the number of User Terminals (UTs) grows large. In the same context, authors in [10] employ IA as an uplink interference mitigation technique amongst cooperating Base Station (BS) clusters for Rayleigh channels. In the context of small cells, the study in [41] extends [10] by assuming clusters of small cells which dictate the use of a Rician fading channel.

The IA technique has also been investigated in multicarrier systems in different settings [15, 19, 38, 62]. A projection based IA technique including the concepts of signal alignment and channel alignment has been investigated in [19]. The IA technique for an interference network with the multicarrier transmission over parallel sub-channels has been tackled in [62]. The signal alignment for multicarrier code division multiple access (MC-CDMA) in two way relay systems has been studied in [38]. Despite various literature about IA in terrestrial cellular networks, only a few studies have been reported about IA in satellite literature. The feasibility of implementing subspace interference alignment (SIA) in a multibeam satellite system has been studied in [30] and it has been concluded that the SIA applied in the frequency domain is advantageous for multibeam satellites.

IA can be broadly classified into two categories: signal level alignment and signal space alignment [37]. The signal level alignment leads to the tractability to DoF characterization while the signal scale alignment provides an attractive way to realize IA in practice. The signal space can be generated in several ways such as by concatenating time symbols, frequency bins, or space domain. Several IA techniques have been reported in the literature based on the availability of CSI knowledge at the transmitter (CSIT), number of signal dimensions used for aligning the interference, and interference removal methods applied at the desired receiver. Existing IA techniques are listed in Table 2.1 along with the corresponding references and briefly described in the following paragraphs.

Linear Interference Alignment: Linear IA is the simplest form of IA in which the alignment of signal spaces is done based on linear precoding (beamforming) schemes. This IA scheme operates within the spatial dimensions provided by multiple antennas at the transmit and receive nodes. Since beamforming schemes are common in the existing point to point MIMO, BC and multiple access networks, linear IA seems to be the most easily accessible form of IA from practical point of view. A linear IA problem becomes a proper or improper based on whether or not the number of equations exceeds the number of variables [75]. The proper systems are likely to be practically feasible and improper systems are likely to be infeasible. Let us consider K user MIMO interference setting with M number of antennas in each transmitter and N number of antennas in each receiver. According [75], the $(M \times N, d)^K$ linear IA problem, d being the number of independent streams, is proper if and only if the following condition is satisfied: $d \leq \frac{M+N}{K+1}$.

Table 2.1 Lists of IA techniques

IA	CSIT	Signal dimensions	Interference removal	References
Linear IA	Perfect/delayed	Single	Filtering	[37, 51, 75]
Subspace IA	Perfect	Multi	Filtering	[37, 64, 70]
Distributed IA	Local	Single	Filtering	[21, 22, 52, 53]
Blind IA	No	Single	Filtering	[23, 26]
Ergodic IA	Perfect/delayed	Single	Filtering	[33, 42]
Asymptotic IA	Perfect	Single	Filtering	[23]
Retrospective IA	Delayed	Single	Filtering	[33, 35, 39]
Lattice alignment	Perfect	Single	Decoding	[4, 44]
Symbol extensions	Multiple channel uses	Fractional	Filtering	[28, 72]
IA and cancelation	Perfect	Single/multi	Filtering and decoding	[73, 74]
Opportunistic IA	Perfect	Single/multi	Filtering	[43, 47]
Asymmetric complex signalling	Complex	Two	Filtering	[8, 29]

Subspace IA: In this scheme, the interferences are aligned to multidimensional subspace instead of a single dimension. In the context of cellular networks, IA scheme provides advantage due to multiuser gain and aligning interferences becomes challenging in the three cell case since there exist multiple non-intended receivers [64]. The IA for one receiver does not guarantee the alignment in the other receivers as well. In fact, this problem arises due to the strict constraint that interferences are mainly aligned into a single dimension. This can be addressed by relaxing the constraints and aligning interferences into multidimensional subspace instead of a single dimension, called as subspace IA. The main concept behind the subspace IA is to align K interfering vectors into $\sqrt{K} + 1$ dimensions (instead of one dimension) to enable simultaneous alignments at the multiple receivers. Since $\sqrt{K}$ becomes negligible compared to K as K gets large, the interference-free DoF can be approached. The interference-free DoF can be achieved as the number of mobiles in each cell i.e., K increases in the context of cellular networks while using the subspace IA. For the G-cell case with K users in each cell, the achievable DoF per cell has been shown to be [64]

$$\frac{K}{(^{G-1}\sqrt{K} + 1)^{G-1}} \to 1 \ \text{ as } \ K \to \infty. \qquad (2.5)$$

Distributed IA: Distributed IA is based on the local channel knowledge instead of global channel knowledge. Several iterative algorithms in the literature have focused on finding the alignment solutions numerically. The motivation for an iterative approach in [22] is to achieve IA with only local channel knowledge, by exploiting the two way nature of communication and the reciprocal nature of the

physical propagation medium. The alternating minimization approach proposed in [49] uses similar distributed IA but does not explicitly assume channel reciprocity. An alternative approach based on weighted minimum mean square error (MMSE) beamforming proposed in [52] compares favorably to the max-Signal to Interference and Noise Ratio (SINR) algorithm and can also provide unequal priorities for the users' rates.

Let us consider a cellular system with B BSs equipped with N antennas and each BS exclusively provides wireless service to K users each equipped with M antennas. The DoFs of a $((B, N) \times (K, 1))$ cellular system with $B > 1$ is given by [53] $d = BK + \frac{BKN}{K+N}$ if N/K is an integer.

Blind IA: Most of the IA results are based on the assumption of perfect, and sometimes, global channel state information at the transmitters. It has been noted that the DoFs of many networks collapse entirely to what is achievable simply by orthogonal TDD among users in the absence of channel knowledge. In this context, there is still possibility of aligning interference based on the knowledge of the distinct autocorrelation properties of the channels observed by different receivers without knowing exact channel coefficients [26]. This is referred as a blind IA technique.

Ergodic IA: In ergodic settings, the channel states can be partitioned into complimentary pairings for a broad class of channel distributions over which the interference can be aligned so that each user is able to achieve (slightly more than) half of his interference-free ergodic capacity at any SNR [42]. The main concept behind this lies on the pairing of channels i.e., matching almost every channel matrix with its complement. Ergodic alignment achieves the capacity when the channel is in a bottleneck state i.e., the number of transmit-receive pairs approaches infinity. In this scheme, each user can achieve at least half of its interference-free capacity at any SNR [42], i.e., $R_k = \frac{1}{2}\mathbb{E}[\log(1 + 2|h_{kk}|^2 P_k)] > \frac{1}{2}R_k^{\text{free}}$, where P_k denotes the transmit power of the kth user, and R_k^{free} denotes the interference-free capacity.

Asymptotic IA: Ergodic IA is an opportunistic scheme that exploits the existence of complementary channel states in equal proportions to achieve the linear IA. Although this assumption applies to a broad variety of channel distributions including Rayleigh fading models, it is not universally applicable since the arbitrary channel distributions, or even standard ones such as Rician fading, do not satisfy the symmetric phase assumptions made by ergodic IA [7]. Although this scheme is of theoretical in nature, it has many advantages such as flexibility of large number of alignment constraints, applicable to both linear and nonlinear forms and for a variety of scenarios ranging from K-user ICs, X networks, cellular networks, compound BC channels, and network coding applications.

Retrospective IA: Retrospective IA techniques refer to the IA schemes that exploit only delayed CSIT. The delayed CSIT is generally assumed to be independent of the current channel state. However, perfect knowledge of channel states is available at the transmitter with some delay. For retrospective IA, the channels can (but do

not have to) be independent and identically distributed (i.i.d.) isotropic [35]. In the absence of the delayed CSIT, i.i.d. isotropic fading channels would lose all signal multiplexing benefits and only have 1 DoF. The result obtained in [35] in the context of a vector BC channel is that CSIT is helpful even if it is outdated and it can have a significant impact since it is capable of increasing the DoF. The delayed feedback can be basically obtained in the following three settings: (i) delayed CSIT: only the past channel states are fed back and not the output signals, (ii) delayed output feedback: only the past received signals are fed back and not the channel states, and (iii) delayed Shannon feedback: the past received signals as well as channel states are fed back. This is the strongest delayed feedback setting, i.e., it can be weakened to obtain either delayed CSIT or the delayed output feedback model by discarding some of the feedback information.

Lattice alignment: Lattice alignment refers to the use of lattice codes in an interference network with the lattices scaled in such a manner that the undesired signals at an interfered receiver arrive on the same lattice, and the desired signal stands apart, i.e., does not occupy the same lattice [4]. The main concept behind this IA scheme is that since the sum of lattice points (codewords) is also a lattice point (a valid codeword), it may be possible to decode the sum of lattice points even if the individual latices by themselves are not decodable. This scheme is mainly applicable for constant channels. Reference [44] considers lattice IA approach for a static real K-user interference channel and derives an achievable rate region for such channels which is valid for finite SNR. For such channels, many results demonstrate that the number of DoFs is very sensitive to slight variations in the direct channel gains.

IA based on Symbol Extensions: Spatial beamforming based linear IA techniques basically operate in the spatial dimensions provided by multiple antennas at the transmit and receive nodes, and divide these spatial dimensions into separable subspaces to be occupied by interference and desired signals at each receiver. In the case of insufficient number of antennas, spatial IA schemes do not find a enough vector space to operate. Furthermore, since the number of beams must be an integer, purely spatial beamforming based IA schemes can only achieve an integral number of signal dimensions per message per channel use. In this case, beamforming across multiple channel uses can be an alternative option to increase the total signal space. For example, the size of the total signal space at each node is increased three times using three channel uses. The concept behind the symbolic extensions is to perform beamforming across multiple channel uses. This technique has been successfully applied for X channel [28] and compound MIMO BC channel [72]. The disadvantage of this approach is that symbol extensions over constant channels do not automatically provide the diversity of linear transformations that is needed for linear IA.

Asymmetric Complex Signalling: Due to lack of rotations in the constant channels while using symbol extensions, the alignment of vector spaces is identical at

each receiver thus making IA infeasible. To overcome the disadvantage of symbol extensions in constant channels, the concept of asymmetric complex signalling has been introduced in [29]. Since we usually deal with the complex numbers for channel coefficients, transmitted and received symbols as well as the noise, phase rotations can be exploited to find distinct rotations at each receiver. This can be realized as rotations in two dimensional real-imaginary plane and this is the main concept behind asymmetric complex signalling method [8, 29].

Interference Alignment and Cancelation: The combination of IA and cancelation (IAC) may be applied to the scenarios where neither IA nor cancelation applies alone. It is shown in [74] that the IAC almost doubles the multiplexing gain (i.e., number of concurrent transmissions) of flat fading interference-limited MIMO channels. In the IAC scheme proposed in [73], the messages are first transformed into asymmetric input with structured coding, and then the dimensions occupied by interference on each receiver are minimized with the help of an appropriate alignment and cancelation technique.

Besides the above techniques, the combined alignment techniques such as signal and channel alignment [19], joint signal and interference alignment [16], joint interference and phase alignment [50] have also been investigated in the literature.

2.3   IA in Cognitive Radio Networks

The IA technique can be classified as an underlay CR technique [20] since it deals with interference mitigation towards the primary system in spectral coexistence scenarios. In the context of coexistence of macrocell and the small cells, authors in [11] have applied the IA technique in order to mitigate the interference from small cells towards the macrocell BS. Similarly, the authors in [40] proposed Vandermonde-subspace frequency division multiplexing for the downlink in order to null out the interference of small cells towards primary macro users. In the coexistence of macro/femto networks, authors in [25] have studied a joint opportunistic interference avoidance scheme with Gale-Shapley spectrum sharing based on the interweave paradigm in order to mitigate both tier interferences. In the proposed scheme, femtocells opportunistically communicate over available spectrum with minimal interference to macrocells while the femtocells are assigned orthogonal spectrum resources to avoid intratier interference. Furthermore, authors in [57] study the application of IA technique exploiting the carrier domain for the coexistence of multibeam and monobeam satellites in order to mitigate the interference of multibeam satellite terminals towards the monobeam satellite. Considering the DoF perspective, the Primary User (PU) does not fully utilize the DOF it can achieve and the primary radio resources are underutilized. In other words, there are free DoFs (DOF holes) in the primary radio resources [14]. As an example, a PU with 1 transmit and 1 receive antenna, who transmits 2 symbols every 3 time slots only utilizes 2/3 DoF while the maximum DoF it can get is 1. So, it is possible for the SUs to access the 1/3 DoF to improve the total DoF of the wireless system.

In the context of CR networks, IA techniques can be broadly classified into non-cooperative and cooperative. Several contributions in the literature have investigated an opportunistic IA scheme in non-cooperative scenarios. The ergodic IA can be considered as an opportunistic scheme that exploits the existence of complementary channel states in equal proportions to achieve IA [48]. The primary CR link can be modeled by a single user MIMO channel since it must operate free of any additional interference caused by secondary systems. Then, assuming perfect CSI at both transmit and receive ends, capacity can be achieved by implementing a water filling power allocation scheme over the spatial directions. It can be noted that even if the primary transmitters maximize their transmission rates, some of their spatial directions are unused due to power limitations. These unused spatial dimensions can therefore be reused by another system operating in the same frequency band in an opportunistic way. An opportunistic secondary transmitter can send its own data to its respective receiver by processing its signal in such a way that the interference produced on the primary link impairs only the unused spatial dimensions. Using the above principle, authors in [47] consider the opportunistic IA considering same number of antennas and same power budget on both primary and secondary devices while authors in [48] consider the opportunistic IA with a general framework where devices have different number of antennas. Furthermore, authors in [1] extend the contribution of [48] considering multiple SUs.

In the context of the cooperative IA technique, authors in [24] study the femto-macro coexistence scenario in order to manage the uplink interference caused by the macrocell users at the femtocell BS (FBS). By means of coordination between multiple FBS and the macrocell users, the received signals from macrocell users can be aligned in a lower dimensional subspace at the multiple FBSs simultaneously. Then the remaining DoFs are exploited to improve the performance of the femtocell users. Similarly, the contribution in [45] considers a cooperative approach to address the interference problem in femtocell networks by allowing the FBSs to perform IA cooperatively in order to reduce their mutual interference and improve the overall performance. Given a number of FBSs deployed over an existing macrocell network, a cooperative strategy is proposed in [45], where the mutual interference inside a coalition of FBSs is aligned in a subspace which is orthogonal to each desired signal. The remaining part of the network, which is non-cooperative, contributes with non-aligned interference on each of the receiver's subspaces.

Furthermore, several IA based cognitive schemes have been proposed in [2] in order to exploit the free spatial dimensions left by the PU. In these schemes, the precoding matrices of the SUs are jointly designed so that no interference is generated at the primary receiver. Furthermore, each secondary receiver does not experience any interference from the primary transmission or from the other SUs. The upper bound of the DoF for a SU (with a single transmitter and receiver) with M_1 antennas at the transmitter and N_1 antennas at the receiver operating in the presence of a PU having d_0 active streams has been found to be [14] $d_1 < \min\{(M_1 - d_0)^+, (N_1 - d_0)^+\}$. Subsequently, for the multiple SUs, each with M number of antennas, the achievable DoF has been found as $(M - d_0)^+$. This bound is the best known bound for cognitive systems without user cooperation

[14]. It indicates that each SU can asymptotically access half the DoF holes. In [14], it is shown that each cognitive user can almost get the whole DoF holes by properly designing their beamforming vectors. According to [14], the number of DoF of the secondary network is given by

$$\max_{\mathcal{D}} \sum_{i=1}^{K} d_i = K \min(\frac{1}{2}, 1 - d_0), \tag{2.6}$$

where $\mathcal{D}$ is the DoF region for the cognitive network and K is the number of SUs. Furthermore, partial and full aided IA schemes can be applied based on the cooperation benefits provided to the PUs.

Moreover, the contribution in [36] studies a trade-off between the Opportunistic Resource Allocation (ORA) and IA techniques in OFDMA based techniques. In the ORA method, the system needs to find an appropriate sub-channel for a femtocell user for which this user has a higher received power from its own BS and less interference from the macrocell transmission so that the total sum-rate is maximized. On the other hand, the IA utilizes fading fluctuations in the frequency domain to generate precoding vectors which create interference-free channels [36]. With the help of numerical results, it has been shown in [36] that the system tends to allocate more sub-channels to perform ORA and achieve the highest sum-rate in low SNR regime while more sub-channels to perform IA in high SNR regime.

2.4 Spectral Coexistence

In this section, we present a generic system model for the spectral coexistence of cognitive systems with primary licensed systems, describing the precoding as well as filtering process. We apply a linear IA technique based on precoding and filtering assuming the perfect CSI knowledge at the secondary transmitters.

2.4.1 Generic System Model

Let us consider a spectral coexistence of a primary system and a secondary system, both operating in a normal uplink mode with the primary system as a single-user uplink and the secondary system as a multiuser uplink. For example, the primary system can be a macrocell system or a monobeam satellite system and the secondary system can be a femtocell system or a multibeam satellite system, which will be described in detail in Sect. 2.5. Usually the primary system is already deployed system and the secondary system should not affect the operation of the primary systems. We consider that the Primary Transmitter (PT) has M signalling dimensions (which can be the number of antennas or carriers) and Primary Receiver

(PR), Secondary Transmitter (ST), and Secondary Receiver (SR) have $L = M + 1$ number of signalling dimensions. This means that there is a single unutilized dimension in the primary link. We consider a single PT, N number of STs and the STs are assumed to be able to cooperate and jointly decode the received signals. Furthermore, the STs are assumed to be aware of the CSI towards the PR and in practice, this knowledge can be obtained by applying the methods mentioned in Sect. 2.2.

In addition to the CSI knowledge, the STs and the PR should be aware of a predefined IA vector, let us denote by $\mathbf{v}$, to perform IA. Depending on how $\mathbf{v}$ is calculated, three different IA techniques can be considered, namely, static, uncoordinated, and coordinated. These techniques basically depend on the level of coordination between primary and secondary systems. The concept behind the applied cognitive IA is to employ precoding at the STs so that the received secondary signals at the PR are all aligned across the alignment vector $\mathbf{v}$. In this way, interference can be filtered out by sacrificing one DoF and some part of the desired received energy. However, after filtering the signal is interference free and can be easily decoded using conventional detection techniques. We mention this technique as cognitive IA since the STs have to be aware of the CSI and the vector $\mathbf{v}$ to perform the precoding. On the other hand, the PR needs only to perform filtering adapted to vector $\mathbf{v}$ and no additional awareness or intelligence is required. The received signal at the PR can be written as:

$$\mathbf{y}_1 = \mathbf{H}\mathbf{x} + \sum_{i=1}^{N} \mathbf{F}_i \mathbf{x}_i + \mathbf{z}_1, \qquad (2.7)$$

where $\mathbf{y}_1$ is the $L \times 1$ received symbol vector, $\mathbf{x}$ is the $M \times 1$ transmitted symbol vector from the PT, $\mathbf{x}_i$ is the $L \times 1$ transmitted symbol vector from the ith ST and $\mathbf{z}_1$ is the receiver noise. All inputs $\mathbf{x}, \mathbf{x}_i$ are assumed to be Gaussian and obey the following sum power constraints: $\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] \leq \gamma_{ps} M$ and $\mathbb{E}[\mathbf{x}_i^\dagger \mathbf{x}_i] \leq \gamma_{ss} L$, γ_{ps} being the transmit SNR of the PT and γ_{ss} being the transmit SNR of the ST. The $L \times M$ matrix $\mathbf{H}$ represents the channel gains between the PR and the PT while the $L \times L$ matrix $\mathbf{F}_i$ represents the channel gains between the PR and ith ST.

Let's group all $\mathbf{F}_i$ into a single $L \times NL$ matrix $\mathbf{F} = [\mathbf{F}_1 \ldots \mathbf{F}_N]$ to simplify notations. The received signal at the joint processor of the SRs is

$$\mathbf{y}_2 = \sum_{i=1}^{N} \tilde{\mathbf{F}}_i \mathbf{x}_i + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{z}_2, \qquad (2.8)$$

where $\mathbf{y}_2$ is the $NL \times 1$ received symbol vector and $\mathbf{z}_2$ is the receiver noise. The $NL \times M$ channel matrix $\tilde{\mathbf{H}}$ represents the channel gains between all SRs and the PT while the $NL \times L$ channel matrix $\tilde{\mathbf{F}}_i$ represents the channel gains between all SRs and the ith ST. To simplify notations, we group all $\tilde{\mathbf{F}}_i$ into a single $NL \times NL$ matrix $\tilde{\mathbf{F}} = [\tilde{\mathbf{F}}_1 \ldots \tilde{\mathbf{F}}_N]$.

2.4.2   IA Precoding and Filtering

Let us assume an $L \times 1$ non-zero reference vector $\mathbf{v}$ along which the interference should be aligned. It should be noted that the STs are assumed to know the alignment direction $\mathbf{v}$ and to have perfect CSI knowledge about the channel coefficients $\mathbf{F}_i$ towards the PR. In this context, the following precoding scheme can be employed to align the interference

$$\mathbf{x}_i = \mathbf{w}_i x_i = (\mathbf{F}_i)^{-1} \mathbf{v} v_i x_i, \qquad (2.9)$$

where $\|\mathbf{v}\|^2 = L$ and $\mathbb{E}[\mathbf{x}_i^\dagger \mathbf{x}_i] \leq L\gamma$. The scaling variable v_i is needed to ensure that the input power constraint is not violated for each ST. This precoding results in unit multiplexing gain and is by no means the optimal IA scheme, but it serves as a tractable way of evaluating the IA performance. Following this approach, the cochannel interference can be expressed as:

$$\sum_{i=1}^{N} \mathbf{F}_i \mathbf{x}_i = \sum_{i=1}^{N} \mathbf{F}_i (\mathbf{F}_i)^{-1} \mathbf{v} v_i x_i = \mathbf{v} \sum_{i=1}^{N} v_i x_i. \qquad (2.10)$$

It can be easily seen that interference has been aligned across the reference vector and it can be removed using an $M \times L$ zero-forcing filter $\mathbf{Q}$ designed in such a way that $\mathbf{Q}$ is a truncated unitary matrix [7] and $\mathbf{Q}\mathbf{v} = \mathbf{0}$. After filtering, the $M \times 1$ received signal vector at the PR can be expressed as:

$$\bar{\mathbf{y}}_1 = \bar{\mathbf{H}}\mathbf{x} + \bar{\mathbf{z}}_1, \qquad (2.11)$$

where $\bar{\mathbf{H}} = \mathbf{Q}\mathbf{H}$ is the $M \times M$ filtered channel matrix. The received signal at the joint processor of the SRs can be written as:

$$\bar{\mathbf{y}}_2 = \sum_{i=1}^{N} \bar{\mathbf{F}}_i x_i + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{z}_2, \qquad (2.12)$$

where $\bar{\mathbf{F}}_i = \tilde{\mathbf{F}}_i (\mathbf{F}_i)^{-1} \mathbf{v} v_i$ are the equivalent $NL \times 1$ channel matrices including precoding. To simplify notations, we group all $\bar{\mathbf{F}}_i$ into a single $NL \times N$ matrix $\bar{\mathbf{F}} = [\bar{\mathbf{F}}_1 \ldots \bar{\mathbf{F}}_N]$. In the following paragraphs, we describe three different IA approaches. The detailed mathematical formulations of these techniques and the theoretical proof that the coordinated approach can perfectly protect the primary rate can be found in [57].

2.4.2.1   Static Approach

In this approach, $\mathbf{v}$ is predefined and does not depend on the channel state. It can be noted that this is quite static but also a simple solution which assumes no

coordination in the network. The disadvantage is that a large amount of received power may be filtered out since the IA direction may be aligned with one of the strong eigenvectors of the random PR-PT channel.

2.4.2.2 Uncoordinated Approach

This approach assumes that the primary and the secondary systems do not coordinate. Furthermore, the STs are aware of their CSI towards the PR but have no information about the CSI of the PT. In this context, the STs select **v** in order to maximize the secondary throughput. Subsequently, the PR senses the **v** and applies the appropriate filter **Q**.

2.4.2.3 Coordinated Approach

In this approach, the primary and secondary systems coordinate to exchange the CSI and the alignment vector. The selection of **v** takes place at the PR and is subsequently communicated to the STs. It is assumed that the channel coherence time is adequate for the alignment direction to be fed back and used by the STs. This is an egoistic approach since the PR dictates the behavior of the STs in order to maximize the performance of the primary system. The coordinated approach perfectly protects the primary rate as reflected in numerical results in Sect. 2.5.

In order to evaluate the system performance of the above techniques, the following two different metrics are considered. The sum-rate capacity of the considered coexistence system is dictated by the primary throughput and the secondary average per-link throughput, let us denote by C_{sys} and define as

$$C_{sys} = C_{ps} + \frac{C_{ss}}{N}, \tag{2.13}$$

where C_{ps} is the throughput of the primary system in the presence of the secondary system, C_{ss} is the average per-link rate of the secondary system in the presence of the primary system, and N is the number of SUs. It should be noted that in (2.13), we consider secondary average per-link throughput i.e., $\frac{C_{ss}}{N}$ in order to reflect the secondary per-user throughput as we increase the number of SUs in the system, as illustrated with the help of numerical results in Sects. 5.1 and 5.2. Subsequently, the primary rate protection ratio is denoted by PR and defined as:

$$PR = \frac{C_{ps}}{C_{po}}, \tag{2.14}$$

where C_{po} denotes the primary only capacity in the absence of the secondary system.
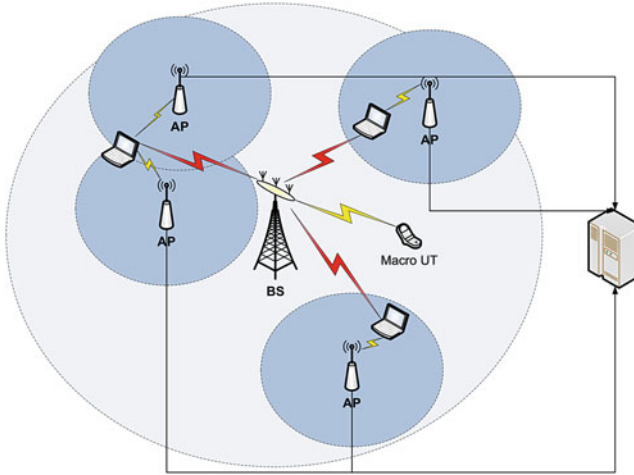
Fig. 2.2 Spectral coexistence scenario of femtocells (secondary) and a macrocell (primary) system using IA

2.5 Practical Scenarios

In this section, we mention two important applications of the IA technique in terrestrial and satellite paradigms based on the generic system and signal models presented in Sect. 2.4. Although these two systems have different characteristics and channel models, they can be studied using the same input-output equations. Furthermore, both systems operate in a normal uplink mode with the primary system as a single user uplink and the secondary system as a multiuser uplink. The only difference between the considered satellite and terrestrial models is that in the terrestrial scenario, IA is over the spatial dimensions and in the satellite scenario, IA is over the subcarriers.

2.5.1 Macrocell-Femtocell Coexistence in Spatial Domain

Let us consider a coexistence scenario of a macrocell and a femtocell systems, both operating in normal uplink mode as shown in Fig. 2.2. The femtocell UTs are STs, femtocell access points (APs) are the SRs, a macro UT is the PT and a macro BS is the PR. Let us consider a coverage area where a single macrocell operates receiving signals from a set of PUs. A number of femtocells (N) operate over the same coverage area receiving signals from a set of SUs. Furthermore, the femtocells are able of cooperating through a broadband backhaul and jointly decoding the received signals. After scheduling, we consider that for a single slot one macro UT and N femtocell UTs are transmitting simultaneously over a common set of frequencies.

Since the macrocell system is the primary, interference coming from the femtocell UTs has to be suppressed. On the other hand, the interference of the macro UT towards the femtocell APs has to be tolerated as the small cell system is secondary. We consider that the macro UT has M antennas while the BS, small cell UTs and the AP have $L = M + 1$ antennas. Furthermore, it is assumed that the interference caused by the small cell UTs have CSI towards the macro BS and this can be easily measured by listening to the macrocell pilot signals.

The considered channel model is based on a MIMO Rayleigh channel whose power is scaled according to a power-law path loss model i.e., asymmetric power levels. More specifically,

$$\mathbf{H} = \alpha\mathbf{G}, \tag{2.15}$$

where α is the path loss coefficient between the BS and the macro UT and $\mathbf{G}$ is an $L \times M$ random matrix with complex circularly symmetric (c.c.s.) i.i.d. elements representing Rayleigh fading coefficients.

The performance of three different IA approaches mentioned in Sect. 2.4 have been compared with the resource division and no-mitigation techniques in [11,57]. Based on the simulation parameters and environment considered in [57], Fig. 2.3 presents the normalized system rate (C_{sys}) versus number of femtocells (N) for the terrestrial coexistence scenario of femtocells and a macrocell. While simulating this scenario, a macro UT and femtocell UTs are considered to be uniformly distributed within the coverage area of the BS and the APs respectively. From the figure (Fig. 2.3), it can be depicted that the sum-rate slowly increases with the value of N for all the considered techniques. The no-mitigation scheme achieves a three-fold gain while other techniques achieve a two-fold gain compared to primary only transmission, however this technique does not protect the primary rate as reflected in Fig. 2.4. Figure 2.4 shows the primary rate protection ratio versus N plots for different techniques. It can be noted that the coordinated IA technique fully protects the primary rate as expected, while other IA techniques preserve roughly 70 % and the resource division preserves 82 % of the primary rate. Furthermore, all techniques except no-mitigation preserve a constant protection rate with increasing N, while the performance of no-mitigation technique degrades monotonically.

2.5.2 Multibeam-Monobeam Satellite Coexistence in Frequency Domain

Recent contributions exploiting spectrum sharing opportunities in satellite communications include [32, 54–59, 61, 67, 69, 76]. The existing cognitive SatComs literature can be categorized into the following: (i) hybrid satellite-terrestrial coexistence scenario [32,54,56,58,59,76] and (ii) dual satellite coexistence scenario [55,57,60,67]. In this section, we present a dual coexistence scenario consisting of
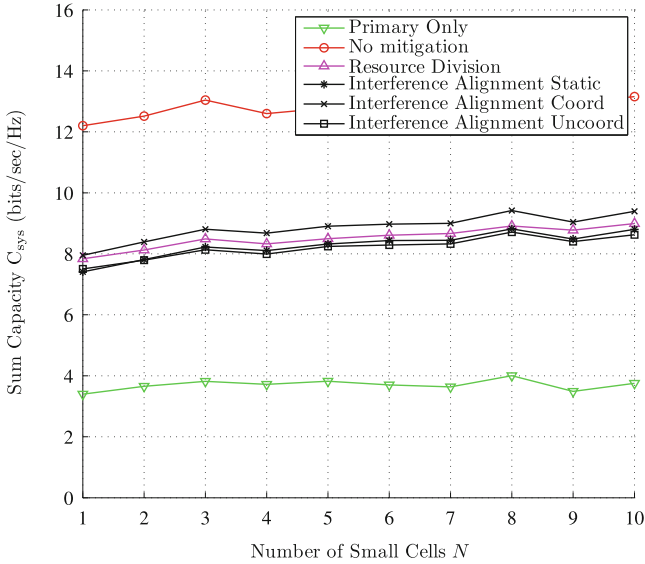
Fig. 2.3 Performance comparison of different techniques in terms of the normalized system rate versus number of small cells N in the considered terrestrial coexistence paradigm
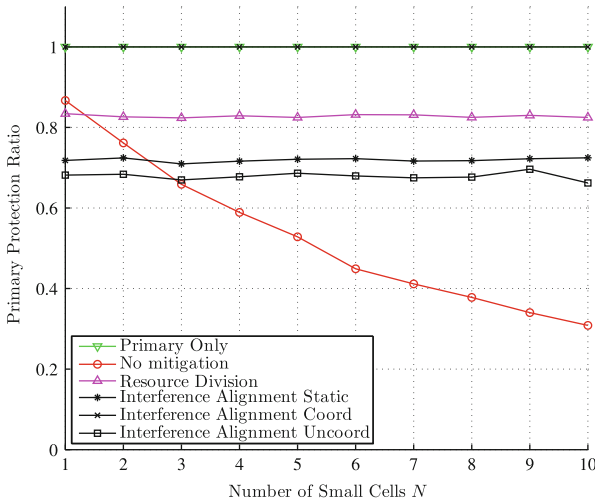
Fig. 2.4 Performance comparison of different techniques in terms of the primary protection ratio versus number of small cells N in the considered terrestrial coexistence paradigm

two multibeam satellites using the IA technique in order to mitigate the interference of multibeam satellite terminals towards the monobeam satellite.

Let us consider one monobeam satellite (SAT1) and one multibeam satellite (SAT2) covering the same area as shown in Fig. 2.5. It can be assumed that they
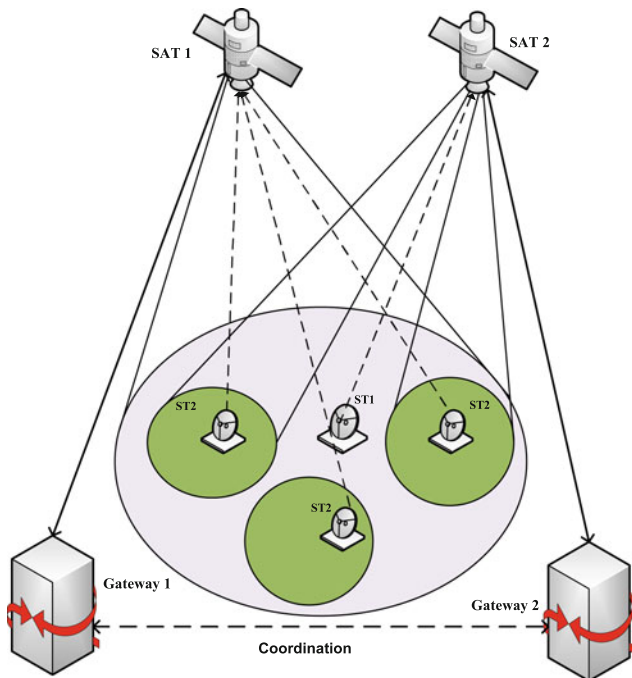
Fig. 2.5 Spectral coexistence scenario of a monobeam satellite (primary) and a multibeam satellite (secondary)

communicate with different gateways on the surface of the Earth. The monobeam satellite uses a single beam to provide coverage to the given area, whereas the multibeam satellite uses several beams to provide coverage to the same area. From the perspective of spectral coexistence, we consider the monobeam system as the primary and the multibeam system as the secondary i.e., the monobeam satellite SAT1 is the PR, the feeders of multibeam satellite SAT2 are the SRs, the multibeam satellite terminals ST2s are the STs and the monobeam satellite terminal ST1 is the PT. In this aspect, the multibeam satellite has to tolerate the interference coming from the monobeam satellite terminal. However, the interference coming from multibeam satellite terminals towards the monobeam satellite has to be suppressed. In this aspect, the IA technique can be applied at the multibeam satellite terminals to mitigate the interference towards the primary satellite.

We consider a single ST1, N number of ST2s served by N beams of SAT2. Multibeam joint processing is considered at the gateway of SAT2 to decode the received signals from ST2s jointly. Since a single gateway is responsible for processing the transmitted and received signals corresponding to a large geographic area, the application of joint processing techniques in the satellite context is centralized. After scheduling, we consider that one ST1 and N number of ST2s

are transmitting simultaneously in a single slot over a common spectrum band. In this context, we consider spatial multiplexing for the primary monobeam system and we employ multiple dimensions (carriers) in the secondary multibeam system to align interference with the reference vector. Furthermore, we consider that all the satellite terminals use multicarrier transmission scheme and the IA is employed at the ST2s over $L = M + 1$ carriers, affected by Adjacent Carrier Interference (ACI). We consider that M number of symbols are transmitted by the ST1 and 1 symbol per ST2 is transmitted by spreading across all the carriers. Furthermore, it should be noted that ST1 sends M symbols over M subcarriers whereas each ST2 sends 1 symbol over L subcarriers. To suppress the interference caused by ST2s using the IA technique, CSI towards the SAT1 is required and we assume that this CSI can be acquired at the ST2s by listening to the pilot signals broadcasted from the gateway.

Each transmit/receive node consists of a single antenna and uses multicarrier transmission so that the channels can be represented as diagonal matrices, where the diagonal entries correspond to different sub-channels. The multicarrier model considered in this scenario differs from MIMO (spatial) channel matrix with full entries as considered in the terrestrial scenario. Due to imperfect bandpass filters, weak copies of adjacent carrier signals may leak into the central carrier causing ACI. Therefore, we consider a multicarrier channel model with ACI. We assume that each carrier goes through independent flat-fading channels. The multi-carrier channel matrix with ACI for the ith satellite link for L number of carriers can be written as

$$\mathbf{H} = \begin{bmatrix} h_1 & \sqrt{\rho}h_2 & \ldots & 0 \\ \sqrt{\rho}h_1 & h_2 & \ldots & 0 \\ 0 & \sqrt{\rho}h_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & h_{L-1} & \sqrt{\rho}h_L \\ 0 & 0 & \sqrt{\rho}h_{L-1} & h_L \end{bmatrix}, \tag{2.16}$$

where ρ represents the fraction of carrier power leaked to adjacent carriers and the parameter h_i represents the Rician fading coefficient, given by;

$$h_i = \left(\sqrt{\frac{R}{R+1}} l + \sqrt{\frac{1}{R+1}} g_i \right), \tag{2.17}$$

where R is the Rician factor, l is a deterministic parameter representing the Line of Sight (LoS) component and g_i is a c.c.s. i.i.d. element for the i-th satellite link representing the Rayleigh fading coefficient.

In the considered satellite coexistence scenario with the simulation parameters in [57], Fig. 2.6 depicts the normalized system rate (C_{sys}) versus number of SAT2 beams N for different techniques and it can be observed that the coordinated

Fig. 2.6 Performance
comparison of different
techniques in terms of the
normalized system rate versus
number of SAT2 beams N in
the considered satellite
coexistence paradigm
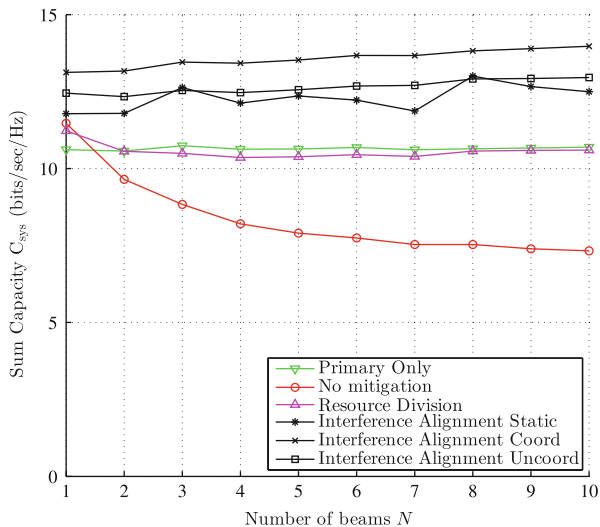
IA technique performs better than all other techniques and the sum-rate slowly
increases with N for this technique. The sum-rate for uncoordinated IA technique
is worse than the coordinated IA technique and is still better than other considered
techniques and it increases slowly with the value of N. Furthermore, the sum-rate
for no mitigation technique decreases with the value of N, remains more or less
constant with the value of N for resource division and remains constant for the
primary only transmission. Since this scenario uses different channel i.e., Rician
fading channel, the sum rate results in Fig. 2.6 differ from that of the results Fig. 2.3.
It should be noted that in the considered satellite coexistence scenario, we use a non-
zero mean channel and consider a tridiagonal channel matrix with three correlated
entries.

Figure 2.7 depicts the PR versus N plot for different techniques. It can be
observed that the coordinated IA technique is optimal and matches with the
primary only technique. This means that the coordinated IA technique fully protects
the primary rate. Furthermore, all techniques except the no-mitigation technique
shows a constant protection rate with the value of N, while the performance
of no-mitigation decreases monotonically as in previous scenario. Moreover, the
uncoordinated IA technique protects almost 90 % of the total primary rate and the
resource division protects about 65 % of the total primary rate. To enhance the
spectrum efficiency further, authors in [13] investigate the effect of the frequency
packing in IA based dual satellite cognitive systems and show that the sum-rate
increases with the value of frequency packing factor for all the considered IA
techniques with the IA coordinated technique perfectly protecting the primary rate
for all the considered frequency packing factor (from 0.5 to 1).

Fig. 2.7 Performance
comparison of different
techniques in terms of the
primary protection ratio
versus number of SAT2
beams N in the considered
satellite coexistence paradigm
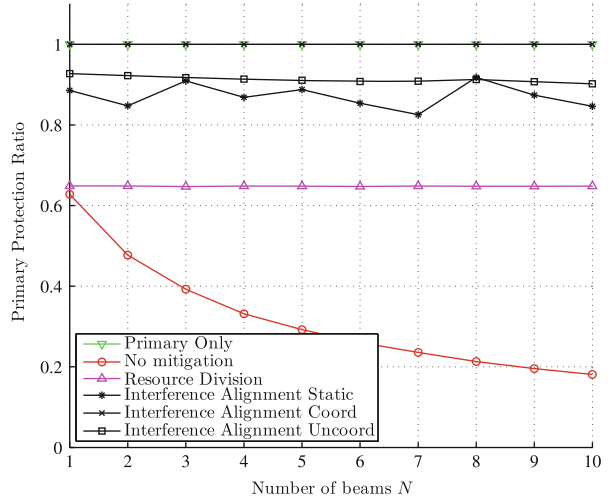
2.6   Practical Challenges of IA

The main practical limitations of IA techniques are requirement of large dimen-
sionality of interference networks, very high SNR, CSI knowledge and strict
synchronization [17]. Other challenges include the overhead for acquiring enough
channel knowledge, the penalty of residual channel uncertainty at the transmitters,
impact of channel correlations, tracking the IA solution under time varying chan-
nels. The main challenge of IA from practical perspectives are listed below.

- In many cases except in the distributed IA, the global channel knowledge is
  required to carry out the IA operation and in distributed IA techniques, local
  CSI knowledge is needed. It's a crucial aspect to investigate suitable blind and
  semiblind IA techniques so that the burden for acquiring the channel knowledge
  is reduced.
- The number of alignment constraints grows very rapidly as the number of
  interfering users is increased. For example, in a K user interference channel, each
  of the K receivers needs an alignment of $K - 1$ interfering signal spaces for a
  total of $O(K^2)$ signal space alignment constraints. Since there are only K signal
  subspaces are available to satisfy $O(K^2)$ signal space alignment constraints, the
  problem can become infeasible [27].
- Limited diversity of interference channels may also limit the relativity of the
  alignment. For example, when each node has only one antenna and all channels
  are constant across time and frequency, the diversity of channels becomes
  limited.
- The practical achievable scheme which requires finite dimensions for the case
  of multiple non-intended receivers is still an open research problem. In this
  context, exploring innovative methods for the optimization of linear precoders

and alignment filters in order to maximize the sum-rate in low and moderate SNR regions is an important future research issue.

2.7 Chapter Summary

Cognitive IA technique could be an effective technique in order allow the coexistence of different wireless networks. In this context, this chapter provides an overview of existing IA techniques along with the principle of IA and the concept of the DoF metric. Furthermore, various existing IA approaches in the context of CR networks have been briefly discussed. Moreover, a general framework for spectral coexistence of wireless networks have been presented and interference alignment and filtering processes have been explained. In addition, two practical coexistence scenarios in satellite and terrestrial paradigms have been illustrated with the help of theoretical and numerical analysis. In addition, several practical challenges of this technique have been identified in order to enable the future research in this domain.

References

1. Abdelhamid, B., ElSabrouty, M., Elramly, S.: Novel interference alignment in multi-secondary users cognitive radio system. In: 2012 IEEE Symposium on Computers and Communications (ISCC), Cappadocia, pp. 000,785–000,789 (2012)
2. Amir, M., El-Keyi, A., Nafie, M.: Constrained interference alignment and the spatial degrees of freedom of MIMO cognitive networks. IEEE Trans. Inf. Theory **57**(5), 2994–3004 (2011)
3. Badic, B., O'Farrrell, T., Loskot, P., He, J.: Energy efficient radio access architectures for green radio: large versus small cell size deployment. In: IEEE 70th Vehicular Technology Conference, Anchorage, pp. 1–5 (2009)
4. Bresler, G., Parekh, A., Tse, D.: The approximate capacity of the many-to-one and one-to-many Gaussian interference channels. IEEE Trans. Inf. Theory **56**(9), 4566–4592 (2010)
5. Cadambe, V., Jafar, S.: Interference alignment and degrees of freedom of the K-user interference channel. IEEE Trans. Inf. Theory **54**(8), 3425–3441 (2008)
6. Cadambe, V., Jafar, S.: Interference alignment and spatial degrees of freedom for the K user interference channel. In: IEEE International Conference on Communications, Beijing, pp. 971–975 (2008)
7. Cadambe, V., Jafar, S.: Interference alignment and the degrees of freedom of wireless X networks. IEEE Trans. Inf. Theory **55**(9), 3893–3908 (2009)
8. Cadambe, V., Jafar, S., Wang, C.: Interference alignment with asymmetric complex signaling–settling the Høst-Madsen-Nosratinia conjecture. IEEE Trans. Inf. Theory **56**(9), 4552–4565 (2010)

9. Chatzinotas, S., Ottersten, B.: Interference alignment for clustered multicell joint decoding. In: IEEE Wireless Communications and Networking Conference, Cancun, pp. 1966–1971 (2011)

10. Chatzinotas, S., Ottersten, B.: Interference mitigation techniques for clustered multicell joint decoding systems. EURASIP J. Wirel. Commun. Netw. Spec. Issue Multicell Coop. Next Gener. Commun. Syst. **2011**(1), 132 (2011)

11. Chatzinotas, S., Ottersten, B.: Cognitive interference alignment between small cells and a macrocell. In: 19th International Conference on Telecommunications, Jounieh, pp. 1–6 (2012)

12. Chatzinotas, S., Zheng, G., Ottersten, B.: Joint precoding with flexible power constraints in multibeam satellite systems. In: IEEE Global Telecommunications Conference, Houston, pp. 1–5 (2011)

13. Chatzinotas, S., Sharma, S.K., Ottersten, B.: Frequency packing for interference alignment-based cognitive dual satellite systems. In: IEEE Vehicular Technology Conference Fall, Las Vegas (2013)

14. Chen, G., Xiang, Z., Xu, C., Tao, M.: On degrees of freedom of cognitive networks with user cooperation. IEEE Wirel. Commun. Lett. **1**(6), 617–620 (2012)

15. Da, B., Zhang, R.: Exploiting interference alignment in multi-cell cooperative OFDMA resource allocation. In: IEEE Global Telecommunications Conference, Houston, pp. 1–5 (2011)

16. Du, H., Ratnarajah, T.: Robust joint signal and interference alignment for MIMO cognitive radio network. In: 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, pp. 448–452 (2012)

17. El Ayach, O., Peters, S., Heath R.W., J.: The practical challenges of interference alignment. IEEE Newblock Wirel. Commun. **20**(1), 35–42 (2013)

18. Etkin, R., Tse, D., Wang, H.: Gaussian interference channel capacity to within one bit. IEEE Trans. Inf. Theory **54**(12), 5534–5562 (2008)

19. Ganesan, R., Klein, A.: Projection based space-frequency interference alignment in a multi-carrier multi-user two-way relay network. In: 8th International Symposium on Wireless Communication Systems, Aachen, pp. 266 –270 (2011)

20. Goldsmith, A., Jafar, S., Maric, I., Srinivasa, S.: Breaking spectrum gridlock with cognitive radios: an information theoretic perspective. Proc. IEEE **97**(5), 894–914 (2009)

21. Gomadam, K., Cadambe, V., Jafar, S.: Approaching the capacity of wireless networks through distributed interference alignment. In: IEEE Global Telecommunications Conference, New Orleans, pp. 1 –6 (2008)

22. Gomadam, K., Cadambe, V., Jafar, S.: A distributed numerical approach to interference alignment and applications to wireless interference networks. IEEE Trans. Inf. Theory **57**(6), 3309–3322 (2011)

23. Gou, T., Wang, C., Jafar, S.: Aiming perfectly in the dark-blind interference alignment through staggered antenna switching. IEEE Trans. Signal Process. **59**(6), 2734–2744 (2011)

24. Guler, B., Yener, A.: Interference alignment for cooperative MIMO femtocell networks. In: 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011), Houston, pp. 1–5 (2011)

25. Huang, L., Zhu, G., Du, X.: Cognitive femtocell networks: an opportunistic spectrum access for future indoor wireless coverage. IEEE Wirel. Commun. **20**(2), 44–51 (2013)

26. Jafar, S.: Exploiting channel correlations – simple interference alignment schemes with no CSIT. In: 2010 IEEE Global Telecommunications Conference (GLOBECOM 2010), Miami, pp. 1–5 (2010)

27. Jafar, S.A.: Interference alignment – a new look at signal dimentions in a communication network. Found. Trends Commun. Inf. Theory **7**(1), 1–134 (2010)

28. Jafar, S., Shamai, S.: Degrees of freedom region of the MIMO X channel. IEEE Trans. Inf. Theory **54**(1), 151–170 (2008)

29. Jafar, S., Cadambe, V., Wang, C.: Interference alignment with asymmetric complex signaling. In: 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton 2009), Piscataway, pp. 991–996 (2009)

30. Jain, P., Vazquez-Castro, M.: Subspace interference alignment for multibeam satellite communications systems. In: 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop, Sardinia, pp. 234–239 (2010)
31. Jung, B.C., Park, D., Shin, W.Y.: Opportunistic interference mitigation achieves optimal degrees-of-freedom in wireless multi-cell uplink networks. IEEE Trans. Commun. **60**(7), 1935–1944 (2012)
32. Kandeepan, S., De Nardis, L., Di Benedetto, M.G., Guidotti A., Corazza G.: Cognitive satellite terrestrial radios. In: IEEE GLOBECOM, Miami, pp. 1–6 (2010)
33. Kang, M., Choi, W.: Ergodic interference alignment with delayed feedback. IEEE Signal Process. Lett. **20**(5), 511–514 (2013)
34. Koo, B., Park, D.: Interference alignment with cooperative primary receiver in cognitive networks. IEEE Commun. Lett. **16**(7), 1072–1075 (2012)
35. Lejosne, Y., Slock, D., Yuan-Wu, Y.: Degrees of freedom in the MISO BC with delayed-CSIT and finite coherence time: optimization of the number of users. In: 2012 6th International Conference on Network Games, Control and Optimization (NetGCooP), Paris, pp. 80–85 (2012)
36. Lertwiram, N., Popovski, P., Sakaguchi, K.: A study of trade-off between opportunistic resource allocation and interference alignment in femtocell scenarios. IEEE Wirel. Commun. Lett. **1**(4), 356–359 (2012)
37. Li, H.: Linear interference alignment based on signal and interference space ranks. In: 4th IET International Conference on Wireless, Mobile Multimedia Networks (ICWMMN 2011), Beijing, pp. 169–172 (2011)
38. Liu, T., Yang, C.: Signal alignment for multicarrier code division multiple user two-way relay systems. IEEE Trans. Wirel. Commun. **10**(11), 3700–3710 (2011)
39. Maleki, H., Jafar, S., Shamai, S.: Retrospective interference alignment over interference networks. IEEE J. Sel. Top. Signal Process. **6**(3), 228–240 (2012)
40. Maso, M., Cardoso, L.S., Debbah, M.: Orthogonal LTE Two-Tier Cellular Networks, Communication (ICC), 2011. IEEE International Conference on, **1**(5), 5–9 (2011)
41. Masucci, A., Tulino, A., Debbah, M.: Asymptotic analysis of uplink interference alignment in ricean small cells. In: IEEE Global Telecommunications Conference, Houston (2011)
42. Nazer, B., Gastpar, M., Jafar, S., Vishwanath, S.: Ergodic interference alignment. IEEE Trans. Inf. Theory **58**(10), 6355–6371 (2012)
43. Nguyen, T.M., Quek, T., Shin, H.: Opportunistic interference alignment in MIMO femtocell networks. In: 2012 IEEE International Symposium on Information Theory Proceedings (ISIT), Cambridge, pp. 2631–2635 (2012)
44. Ordentlich, O., Erez, U.: On the robustness of lattice interference alignment. IEEE Trans. Inf. Theory **59**(5), 2735–2759 (2013)
45. Pantisano, F., Bennis, M., Saad, W., Debbah, M., Latva-aho, M.: Interference alignment for cooperative femtocell networks: a game-theoretic approach. IEEE Trans. Mobile Comput. **12**(11), 2233–2246 (2012)
46. Parker, P.A., Bliss, D., Tarokh, V.: On the degrees-of-freedom of the MIMO interference channel. In: 42nd Annual Conference on Information Sciences and Systems (CISS 2008), Princeton, pp. 62–67 (2008)
47. Perlaza, S., Debbah, M., Lasaulce, S., Chaufray, J.M.: Opportunistic interference alignment in MIMO interference channels. In: IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2008), Cannes, pp. 1–5 (2008)
48. Perlaza, S., Fawaz, N., Lasaulce, S., Debbah, M.: From spectrum pooling to space pooling: opportunistic interference alignment in MIMO cognitive networks. IEEE Trans. Signal Process. **58**(7), 3728–3741 (2010)
49. Peters, S., Heath, R.: Interference alignment via alternating minimization. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, pp. 2445–2448 (2009)
50. Razavi, S., Ratnarajah, T., Masouros, C., Sellathurai, M.: Joint interference and phase alignment in multiuser MIMO interference channels. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, pp. 1137–1141 (2012)

51. Razaviyayn, M., Sanjabi, M., Luo, Z.Q.: Linear transceiver design for interference alignment: complexity and computation. IEEE Trans. Inf. Theory **58**(5), 2896–2910 (2012)
52. Schmidt, D., Shi, C., Berry, R., Honig, M., Utschick, W.: Minimum mean squared error interference alignment. In: 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, pp. 1106–1110 (2009)
53. Schreck, J., Wunder, G.: Distributed interference alignment in cellular systems: Analysis and algorithms. In: 11th European Wireless Conference 2011 – Sustainable Wireless Technologies (European Wireless), Vienna, Austria, pp. 1–8 (2011)
54. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Satellite cognitive communications: interference modeling and techniques selection. In: 6th ASMS/SPSC Conference, Baiona, pp. 111–118 (2012)
55. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Exploiting polarization for spectrum sensing in cognitive SatComs. In: 7th International Conference CROWNCOM, Stockholm, pp. 36–41 (2012)
56. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Spectrum sensing in dual polarized fading channels for cognit SatComs. In: IEEE Globecom Conference, Anaheim, pp. 3443–3448 (2012)
57. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Interference alignment for spectral coexistence of heterogeneous networks. EURASIP J. Wirel. Commun. Netw. **46** (2013). doi:10.1186/1687-1499-2013-46
58. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Transmit beamforming for spectral coexistence of satellite and terrestrial networks. In: 8th International Conference CROWNCOM, Washington, D.C., pp. 275–281 (2013)
59. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Spatial filtering for underlay cognitive SatComs. In: Proceedings of the 5th International Conference PSATS, Toulouse (2013)
60. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Cognitive beamhopping for spectral coexistence of multibeam satellites. In: Future Network Mobile Summit, Lisbon (2013)
61. Sharma, S.K., Chatzinotas, S., Ottersten, B.: Cognitive radio techniques for satellite communication systems. In: IEEE VTC-fall, Las Vegas (2013)
62. Shi, C., Berry, R., Honig, M.: Interference alignment in multi-carrier interference networks. In: IEEE International Symposium on Information Theory Proceedings, Saint Petersburg, pp. 26–30 (2011)
63. Shin, W., Lee, N., Lim, J.B., Shin, C., Jang, K.: On the design of interference alignment scheme for two-cell MIMO interfering broadcast channels. IEEE Trans. Wirel. Commun. **10**(2), 437–442 (2011)
64. Suh, C., Tse, D.: Interference alignment for cellular networks. In: 46th Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champaign, IL, pp. 1037–1044 (2008)
65. Telatar, E.: Capacity of multi-antenna Gaussian channels. Eur. Trans. Telecommun. ETT **10**(6), 585–596 (1999)
66. Tresch, R., Guillaud, M., Riegler, E.: On the achievability of interference alignment in the K-user constant MIMO interference channel. In: IEEE/SP 15th Workshop on Statistical Signal Processing, Cardiff, pp. 277–280 (2009)
67. Vassaki, S., Poulakis, M.I., Panagopoulos, A.D., Constantinou, P.: Power allocation in cognitive satellite terrestrial networks with QoS constraints. IEEE Commun. Lett. **17**(7), 1344–1347 (2013)
68. Vaze, C., Varanasi, M.: The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT. IEEE Trans. Inf. Theory **58**(8), 5354–5374 (2012)
69. Wang. L.N., Wang, B.: Distributed power control for cognitive satellite networks. Adv. Mater. Res. Mechatron. Intell. Mater. II **71**, 1156–1160 (2012)
70. Wang, C., Gou, T., Jafar, S.: Subspace alignment chains and the degrees of freedom of the three-user MIMO interference channel. In: 2012 IEEE International Symposium on Information Theory Proceedings (ISIT), Cambridge, pp. 2471–2475 (2012)

71. Weingarten, H., Steinberg, Y., Shamai, S.: The capacity region of the Gaussian multiple-input multiple-output broadcast channel. IEEE Trans. Inf. Theory **52**(9), 3936–3964 (2006)
72. Weingarten, H., Shamai, S., Kramer, G.: On the compound MIMO broadcast channel. In: Proceedings of Annual Information Theory and Applications Workshop UCSD, San Diego (2007)
73. Yang, L., Zhang, W.: On design of asymmetric interference alignment and cancelation scheme in MIMO X network. In: 2011 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, pp. 1–5 (2011)
74. Yang, L., Zhang, W.: Asymmetric interference alignment and cancelation for 3-user MIMO interference channels. In: 2012 IEEE International Conference on Communications (ICC), Sydney, pp. 2260–2264 (2012)
75. Yetis, C., Gou, T., Jafar, S., Kayran, A.: On feasibility of interference alignment in MIMO interference networks. IEEE Trans. Signal Process. **58**(9), 4771–4782 (2010)
76. Yun, Y.H., Cho, J.H.: An orthogonal cognitive radio for a satellite communication link. In: IEEE 20th International Symposium PIMRC, Tokyo, pp. 3154–3158 (2009)

Chapter 3
Cooperative Spectrum Sensing

H. Birkan Yilmaz, Salim Eryigit, and Tuna Tugcu

Abstract The main objective of this chapter is to provide a detailed technical insight into latest key aspects of cooperative spectrum sensing. We focus on fusion strategies, quantization enhancements, effect of imperfect reporting channel, cooperative spectrum sensing scheduling, and utilizing cooperatively sensed data via Radio Environment Map (REM).

The sharing of local observations between the secondary users and the fusion center is the most crucial factor that determines the performance of cooperative sensing. Detection performance is determined by the quality of local observations and the quality of the information received by the fusion center. Therefore, the number of quantization bins, the number of bits sent for sensing reports, the global decision logic, and the imperfections in the reporting channel and the erroneous reports due to malfunctioning or malicious secondary devices affect the system performance. Furthermore, there are many channels to sense while the cooperating nodes are few, therefore coordinating the sensing nodes for detecting high quality channels is necessary. Cooperative sensing scheduling concentrates on the scheduling of cooperative nodes and the channels to be sensed.

There is an intricate interplay among the period and size of the sensing reports and spectral resources. Decreasing the number of bits for sensing reports with acceptable performance enables increasing the number of sensing periods and better performance. Furthermore, having a bandwidth-limited reporting channel does not allow sending the whole observation and using complicated protocols for sending the sensing reports to the fusion center. Hence, this chapter also focuses on quantization enhancements.

Due to the periodic sensing requirement of a typical cognitive radio network, cumulative energy consumption for sensing becomes a challenging factor. The

H.B. Yilmaz (✉) • S. Eryigit • T. Tugcu
Department of Computer Engineering, Bogazici University, Istanbul, Turkey
e-mail: birkan.yilmaz@boun.edu.tr; eryigit@boun.edu.tr; tugcu@boun.edu.tr

energy problem becomes more severe if the users are mobile. The components of energy consumption dedicated to cooperative sensing are analyzed and optimal, and sub-optimal (but efficient) sensing scheduling mechanisms are discussed in order to reduce the sensing energy consumption of the network.

Once the spectrum has been sensed cooperatively, the outcomes can be utilized via REM, which can be considered as an important part of the cognitive engine located at the network. The sensed information may also play a crucial role in the generation of REM. Hence, this chapter also focuses on how the sensing measurements could be utilized for REM construction.

3.1 Introduction

Cooperative sensing involves the discussion of various topics. This chapter is organized in three main passages to lead the reader step by step in understanding the overall concept. The first passage is composed of the following five sections where spectrum sensing and cooperation strategies are discussed. The second passage builds on the first one and covers how the sensing nodes (typically secondary users – SU) should be assigned to the channels to be sensed in order to maximize the throughput. Finally, the third passage utilizes the experience accumulated in the previous passages and complements them by introducing how the sensing information acquired from all nodes in the field can be transformed into "knowledge" about the spectrum use and availability by constructing a radio environment map (REM).

The first passage starts with the discussion of how sensing can be done under different channel conditions, namely additive white Gaussian noise (AWGN), Rayleigh, and Nakagami-m fading channels. Once the each node in the field has performed the sensing task individually, they should transmit their local measurements to a central decision maker for cooperation. Cooperative sensing can be implemented in this central node by means of different fusion strategies. The local measurements are typically transmitted to the central node after some quantization process to utilize the scarce bandwidth of the reporting channel. Yet, the quantization method employed is one of the important factors that affect the performance of cooperative sensing and is closely related with the fusion strategies. Another factor that may affect the performance is the imperfections in the reporting channel; if some nodes transmit erroneous sensing reports, intentionally or not, some fusion strategies may simply fail. Thus, the detector should be optimized to benefit from quantization while tolerating imperfections in the reporting channel.

Considering the fact that multiple sensing nodes are required to achieve cooperative sensing and there are many channels to be surveyed, once can easily conclude that SUs should be carefully assigned to the channels to be sensed according to their perceived SNRs. Thus in the second passage, measurements from the SUs that are expected to sense the channel better are considered in the cooperative decision. Furthermore, intelligent scheduling of the sensing nodes can help in reducing energy

consumption. Thus, the scheduling system is modeled as a minimization problem and solved for optimality.

Now that the channels are sensed by carefully selected multiple SUs and the measurements are cooperatively evaluated for long durations, we have a huge amount of information about all channels in the field. Yet, it is a difficult but vital task to determine which channel is more likely to be available at which part of the terrain at any moment. The third passage covers how a REM can be constructed from the gathered information and introduces two quality metrics for evaluation. The REM serves as a useful tool for the cognitive engine in a cognitive radio network.

3.2 Spectrum Sensing Preliminaries

Spectrum sensing is the most important task in the cognitive cycle for the realization of cognitive radio. Since cognitive radios are considered lower priority or SU of the spectrum band allocated to a Primary User (PU), a fundamental requirement is to avoid interference to potential PUs in the vicinity. On the other hand, PU networks are not required to change their infrastructure for spectrum sharing with cognitive networks. Therefore, cognitive radio networks should be able to independently detect PU presence through spectrum sensing schemes. Although spectrum sensing is traditionally considered as measuring the spectral content or measuring the interference over the spectrum, it is a more general term that involves obtaining the spectrum usage characteristics across multiple dimensions such as time, space, frequency, and code [1].

In a nutshell, the goal of spectrum sensing is to decide between two hypotheses

$$
\begin{aligned}
Y(t) &= n(t) & H_0 \ (\textit{white space}), \\
Y(t) &= h \times s(t) + n(t) & H_1 \ (\textit{occupied})
\end{aligned}
\tag{3.1}
$$

where $Y(t)$ is the complex signal received by the cognitive radio device, $s(t)$ is the transmitted signal of the primary user, $n(t)$ is the additive white Gaussian noise, and h is the complex gain of the ideal channel.

Common transmitter detection methods in the literature are *Energy Detection (ED)*, *Matched Filter Detection*, and *Cyclostationary Detection*. The ED based approach, also known as radiometry or periodogram, is the most common way of spectrum sensing in high SNR conditions since it does not require any a priori knowledge of primary signals and has much lower computational and implementation complexity [4, 5, 14, 20–22, 28, 39, 43, 47, 51, 62].

The conventional energy detector depicted in Fig. 3.1 uses a single threshold to determine the presence or absence of the signal. The input band-pass filter selects the center frequency, f_c, and the bandwidth of interest in Hz, W. This filter is followed
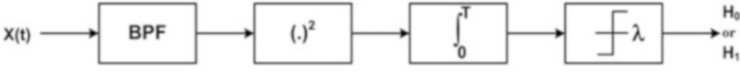
Fig. 3.1 Block diagram of the conventional energy detector

by a squaring device to measure the received energy and an integrator which determines the observation interval in seconds, T. Finally, the output is compared with a threshold, λ, to decide whether the signal is present. The presence of a primary user in AWGN as well as flat fading Rayleigh channels results in different ED outputs and also ED output is affected by the dynamic channel conditions.

3.2.1 Spectrum Sensing in AWGN Channel

Under an AWGN channel, the energy received ($O_i = \sum_{j=1}^{2TW} Y_{ij}^2$) by a secondary user i follows the distribution

$$f(O_i|\gamma) \approx \begin{cases} \chi_{2TW}^2 & H_0 \\ \chi_{2TW}^2(2\gamma) & H_1 \end{cases} \tag{3.2}$$

where χ_{2TW}^2 and $\chi_{2TW}^2(2\gamma)$ represent the central and the non-central chi square distributions [15, 28, 55], TW and γ represent the time bandwidth product and SNR, respectively. Under AWGN channel conditions, the SNR value is fixed and it affects the separation between conditional probability distribution functions. A high SNR separates the distributions enough to decide safely and with a reasonable probability of error. However, under low SNR conditions, it is difficult to distinguish between H_0 and H_1 probability distributions.

In the single threshold ED, the decision for H_0 and H_1 depends solely on λ. The probability of detection, $\mathbf{P_D}^A$, and the probability of false alarm, $\mathbf{P_F}^A$, for a single SU under AWGN channel can be calculated using λ with the exact closed-form equations

$$\mathbf{P_D^A}(\lambda) = \mathbf{P}\{O_i > \lambda \ |H_1\} = \mathbf{Q}_{TW}(\sqrt{2\gamma}, \ \sqrt{\lambda}) \tag{3.3}$$

$$\mathbf{P_F^A}(\lambda) = \mathbf{P}\{O_i > \lambda \ |H_0\} = \frac{\Gamma(TW, \lambda/2)}{\Gamma(TW)} \tag{3.4}$$

where $\mathbf{Q}_u$ is the Marcum Q-Function with degree of freedom u and $\Gamma(.), \Gamma(.,.)$ represent the Gamma and Incomplete Gamma Functions, respectively [15, 55].

3.2.2 Spectrum Sensing in Rayleigh Channel

The mobile radio channel is characterized by the multipath reception. The signal reaching the receiver contains not only a direct line-of-sight radio wave, but also a large number of reflected radio waves [27]. Even worse in urban centers, the line-of-sight is often blocked by obstacles. The basic model of Rayleigh fading assumes the received multipath signal to consist of a large number of reflected waves with i.i.d. inphase and quadrature amplitudes, and not the line-of-sight component. Under the Rayleigh fading channel, the probability of false alarm, $\mathbf{P_F^R}$, remains the same as in the case of AWGN channel since it depends only on the distribution of noise. However, under the Rayleigh channel with no diversity, the signal amplitude follows Rayleigh distribution [15]. Therefore, SNR γ follows the exponential pdf

$$f(\gamma) = \frac{1}{\overline{\gamma}} \exp(-\gamma/\overline{\gamma}), \quad \gamma \geq 0. \tag{3.5}$$

The average $\mathbf{P_D}$ in this case, $\mathbf{P_D^R}$, can be calculated by averaging Eq. 3.3 over Eq. 3.5 [14, 15].

$$\mathbf{P_D^R}(\lambda) = \int_{\gamma} \mathbf{P_D^A} f(\gamma) d\gamma \tag{3.6}$$

$$\mathbf{P_D^R}(\lambda) = e^{-\lambda/2} \sum_{n=0}^{TW-2} \frac{1}{n!} \left(\frac{\lambda}{2}\right)^n$$

$$+ \left(\frac{1+\overline{\gamma}}{\overline{\gamma}}\right)^{TW-1} \left[e^{-\frac{\lambda}{2(1+\overline{\gamma})}} - e^{-\frac{\lambda}{2}} \sum_{n=0}^{TW-2} \frac{1}{n!} \left(\frac{\lambda\overline{\gamma}}{2(1+\overline{\gamma})}\right)^n \right] \tag{3.7}$$

3.2.3 Spectrum Sensing in Nakagami-m Fading Channel

Under the Rayleigh fading channel in which diversity paths are i.i.d., the output SNR, γ_t, is the sum of the SNRs on all branches and can be used for evaluating the average $\mathbf{P_D}$ for EGC scheme.

The pdf of γ_t for i.i.d. Rayleigh branches is given by

$$f(\gamma_t) = \frac{1}{(L-1)!\overline{\gamma}^L} (\gamma_t)^{L-1} \exp(-\gamma_t/\overline{\gamma}) \tag{3.8}$$

where L is the number of i.i.d. diversity branches. The pdf in Eq. 3.8 is similar to the pdf of SNR in the Nakagami channel. The Nakagami parameter m can be viewed as a diversity order. Hence, the average $\mathbf{P_D}$ for the EGC scheme, $\mathbf{P_D}^E$, is obtained by replacing m, $\overline{\gamma}$, and TW by L, $L\overline{\gamma}$, and LTW, respectively [15]. We take the diversity order L equal to the number of cooperating nodes for evaluating $\mathbf{P_D}^E$. Hence, the $\mathbf{P_D}^E$ formulation becomes

$$\mathbf{P_D}^E(\lambda) = \alpha \left[\Psi + \beta \sum_{n=1}^{LTW-1} \frac{(\lambda/2)^n}{2n!} \, {}_1F_1(L; n+1; \frac{\lambda}{2} \frac{\overline{\gamma}}{(1+\overline{\gamma})}) \right] \qquad (3.9)$$

where ${}_1F_1(.;.;.)$ is the confluent hypergeometric function,

$$\alpha = \frac{1}{\Gamma(L)2^{L-1}} \left(\frac{1}{\overline{\gamma}} \right)^L , \qquad (3.10)$$

$$\beta = \Gamma(L) \left(\frac{2\overline{\gamma}}{1+\overline{\gamma}} \right)^L \exp(-\lambda/2), \qquad (3.11)$$

and

$$\Psi = \frac{2^{L-1}(L-1)! \, \overline{\gamma}}{(1/\overline{\gamma})^L (1+\overline{\gamma})} e^{-\frac{\lambda}{2(1+\overline{\gamma})}} \left[(1 + \frac{1}{\overline{\gamma}})(\frac{1}{1+\overline{\gamma}})^{L-1} L_{L-1}(-\frac{\lambda\overline{\gamma}}{2(1+\overline{\gamma})}) + \right.$$
$$\left. \sum_{n=0}^{L-2} (\frac{1}{1+\overline{\gamma}})^n L_n(-\frac{\lambda\overline{\gamma}}{2(1+\overline{\gamma})}) \right] \qquad (3.12)$$

where $L_n(.)$ is the Laguerre polynomial of degree n [15]. Since we do not differentiate the nodes with different SNR levels, EGC constitutes an upper bound for optimization problems considered. These formulations are focusing on local ED, however cooperation increases the performance of the detection process.

3.3 Cooperation and Fusion Strategies

The sensing methods discussed in the previous section are still subject to failure due to channel conditions in different parts of the terrain. Thus, it is possible that some SUs may detect a primary user while some other do not, depending on the location. To take advantage of the spatial diversity in the wireless channel, cooperative spectrum sensing methods have been proposed in [2, 31, 54, 64]. It is shown analytically and through numerical results that cooperative sensing provides

significant higher spectrum capacity gains than local sensing [32]. The detection performance is determined by the quality of local observations and the quality of the information received by the fusion center where the cooperation is achieved. Therefore, the number of quantization bins, the number of bits sent for sensing reports, and the global decision logic affect the overall system performance.

In the case of cooperative sensing, sharing information among CRs and combining results from various measurements is a challenging task. The shared information can be soft or hard decisions made by each cognitive device [56]. The results presented in [56,58] show that the soft information-combining outperforms the hard information-combining method in terms of the probability of missed opportunity. On the other hand, hard-decisions are found to perform as good as soft decisions when the number of cooperating users is high [34]. *Therefore, cooperation and quantization for bandwidth limited reporting channel promise increasing the performance of the detector under a low SNR regime.*

3.3.1 Hard Fusion Strategies

When hard decisions are used, And, Or, Majority, and M-out-of-N methods can be used for combining the information from different cognitive radios [40]. In the And-rule, all sensing results should be H_1 for deciding H_1, where H_1 is the alternate hypothesis, i.e. the hypothesis that the observed band is occupied by a primary user. In the Or-rule, the fusion center decides H_1 if any of the received decisions plus its own is H_1. In the Majority-rule, for deciding H_1 it is required to have majority of the nodes having decision H_1. M-out-of-N rule outputs H_1 when the number of H_1 decisions is equal to or larger then M where there are N cooperating nodes. Therefore, And, Or, and Majority rules may be considered as special cases of M-out-of-N rule.

3.3.2 Soft Fusion Strategies

The optimum fusion rule for combining the sensing information is the Chair-Varshney rule which is based on log-likelihood ratio test [7]. Likelihood ratio tests are used for making classification using decisions from secondary users in [6, 18, 19, 56]. Various simpler techniques for combining sensing results are employed in [14]. The performances of EGC, Selection Combining (SC), and Switch and Stay Combining (SSC) are investigated for energy detector based spectrum sensing under Rayleigh fading. The EGC method is found to have a gain of approximately two orders of magnitude gain while SC and SSC having one order of magnitude gain.

3.4 Quantization Enhancements

Let's consider the cooperative sensing methods that utilize quantization. In [29], cooperative sensing and quantization schemes (1-bit or hard decision) are investigated for multiple primary bands. The spectrum sensing scheduling and sensing time are analyzed in [9, 65]. In [49], the authors analyze cooperative sensing under bandwidth constraints. In [42], the authors examine the optimal quantizer for signal detection locally. The work is extended in [37] by using evidence theory based cooperative spectrum sensing with efficient quantization. Similarly in [2], detection and false alarm probabilities are derived with consideration of errors in the reporting channel due to fading. Only the hard decision logic M-out-of-N is studied in the paper. In [46] and [45], cooperative sensing via quantization is studied for 2-bit and 3-bit quantization. However, the global decision logic is static in terms of weights.

3.4.1 Incentives for Utilizing Quantization

The conventional ED is a quantizer with two bins that uses single threshold to determine the presence or absence of the signal. By using more quantization bins at the sensing nodes, the information gain accrued by cooperation can be increased. Thresholds divide the observation space into bins and the sensing node determines the bin into which the observation falls. Most of the works on quantization focus on just the local quantization process. Therefore, quantization enhancement has two parts: local quantization and global decision logic.

3.4.2 Local Quantization

In Fig. 3.2, quantization levels with four bins are depicted as an example at the bottom of the figure. There are three thresholds that are determined by the first threshold, λ_1, since the distance between consecutive thresholds is fixed and denoted by Δ (i.e., uniform quantizer). The observation of node i, O_i, is greater than or equal to zero, and the observation space is divided into bins where B_k denotes the kth bin.

For a given λ_1, other thresholds can be found by adding Δ for the next threshold at each step. Hence, the degree of freedom is just one for a given Δ. For a given set of thresholds $\{\lambda_1, \lambda_2, \ldots, \lambda_{n-1}\}$, we can evaluate the probability of having observation O_i in B_k under H_0 and H_1, respectively. $\mathbf{P}_{H_i}^A(B_k)$ denotes the probability of having local observation in bin B_k under hypothesis H_i and AWGN channel as
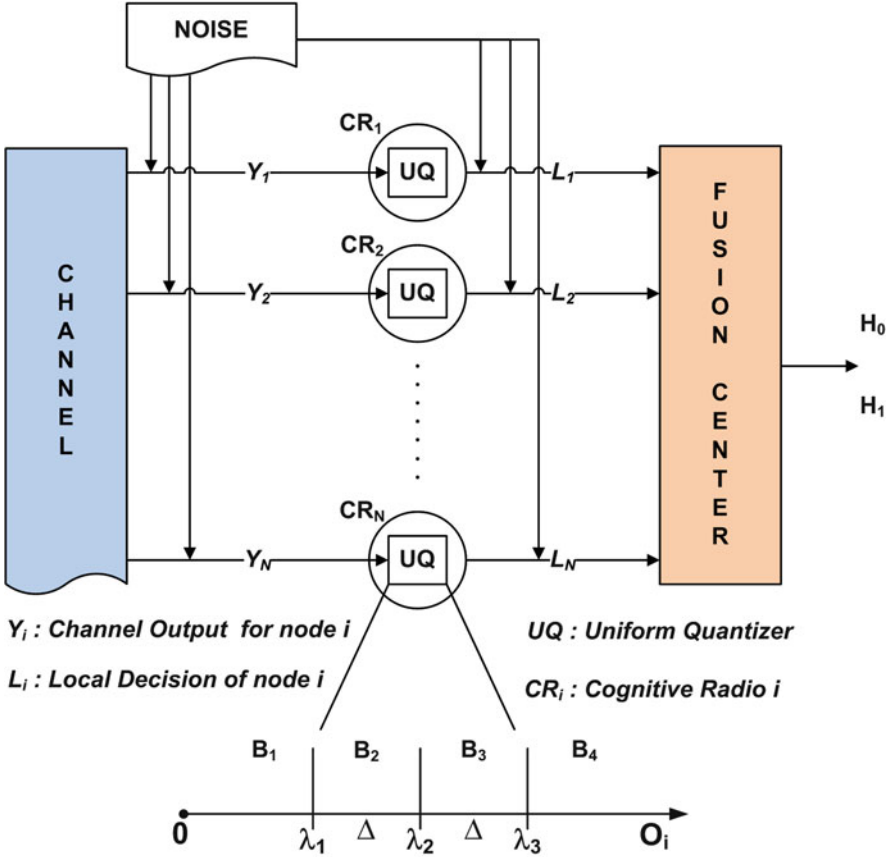
Fig. 3.2 System model of cooperative methods and local quantization

$$\mathbf{P}_{H_0}^{A}(B_k) = \begin{cases} 1 - G_{TW}(\lambda_k) & \text{if } k = 1 \\ G_{TW}(\lambda_{k-1}) & \text{if } k = n \\ G_{TW}(\lambda_{k-1}) - G_{TW}(\lambda_k) & \text{otherwise} \end{cases} \quad (3.13)$$

$$\mathbf{P}_{H_1}^{A}(B_k) = \begin{cases} 1 - Q_{TW}(\gamma, \lambda_k) & \text{if } k = 1 \\ Q_{TW}(\gamma, \lambda_{k-1}) & \text{if } k = n \\ Q_{TW}(\gamma, \lambda_{k-1}) - Q_{TW}(\gamma, \lambda_k) & \text{otherwise} \end{cases} \quad (3.14)$$

where $G_{TW}(\lambda)$ denotes $\frac{\Gamma(TW, \lambda/2)}{\Gamma(TW)}$ and $Q_{TW}(\gamma, \lambda)$ denotes $\mathbf{Q}_{TW}(\sqrt{2\gamma}, \sqrt{\lambda})$.

Under a Rayleigh channel and conditioned on H_0, there is no difference with the AWGN case since there is no transmission. For evaluating bin probabilities conditioned on H_1, Eq. 3.7 of $\mathbf{P}_{\mathbf{D}}^{R}$ is used. A similar bin probability formula can

be easily obtained for a Rayleigh channel by replacing $Q_{TW}(\gamma, \lambda)$ in Eq. 3.14 with the $\mathbf{P_D^R}$ formulation.

$$\mathbf{P}_{H_1}^R(B_k) = \begin{cases} 1 - \mathbf{P_D^R}(\lambda_k) & \text{if } k = 1 \\ \mathbf{P_D^R}(\lambda_{k-1}) & \text{if } k = n \\ \mathbf{P_D^R}(\lambda_{k-1}) - \mathbf{P_D^R}(\lambda_k) & \text{otherwise} \end{cases} \qquad (3.15)$$

3.4.3 Fusion Strategies

To take advantage of the spatial diversity in the wireless channel, cooperative spectrum sensing methods have been proposed in [31, 54, 64]. We provide a cooperative sensing scheme for similar reasons.

The system model for the proposed method is depicted in Fig. 3.2. Each cooperating secondary user senses the spectrum and sends its "quantized" local measurement as L_i, (index of the quantization bin) to the fusion center at the cognitive base station. The fusion center makes a global decision according to L_i values.

The global decision logic schemes in the literature can be classified into two main categories: soft decision and hard decision logic. In hard decision methods, the fusion center collects local decisions consisting of 1-bit information. In soft decision methods, the exact measurements are reported to the fusion center. Well known soft decision methods are EGC and Maximal Gain Combiner (MGC). EGC uses fixed weights for measurements reported to the fusion center. All received measurements are summed coherently and compared against one global threshold. In MGC, the received measurements are weighted with respect to their SNR values, and then summed and compared with one predetermined global threshold.

If the quantization is done using basically two bins, hard decision logic functions, such as Or and Majority Logic, can be applied as the global decision logic at the fusion center. Considering 0 and 1 for L_i trivially suggests hard decision logic for cooperation. However, in the case of more quantization bins none of the hard decision logic functions can be used. In such a situation, the global decision logic must have a functional form.

Yilmaz et al. propose the global decision technique, *Cooperative Sensing with Decision Vector (CSDV)* [61] which is inspired by the analogy of a seesaw. The example in Fig. 3.3 contains three measurements in B_1, one measurement in B_2, and five measurements in B_3, hence the final decision is clearly H_0 with the given weights, even though the majority of the nodes are at the right side.

For the given 4-bin example in Fig. 3.3, the fusion center receives the quantized measurements and counts the number of users in each quantization bin. Having three reports in B_1, one report in B_2, and five reports in B_3 implies $\vec{B} = [3\ 1\ 5\ 0]$. The decision function, $\delta_{\vec{w}}(.)$, is evaluated with the help of the weights and the number of users in the bins as
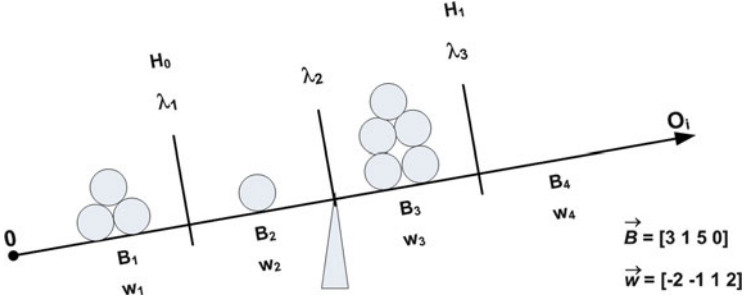
Fig. 3.3 Seesaw analogy of the global decision logic
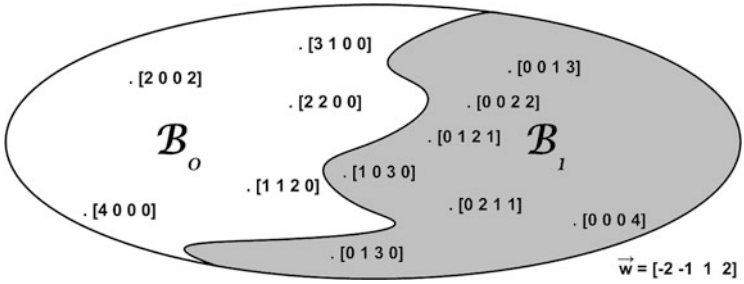
Fig. 3.4 Set of local decisions for four nodes and quantization bins

$$
\delta_{\vec{w}}(\vec{B}) = \begin{cases} 1 & \text{if } \vec{B} \cdot \vec{w} > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3.16}
$$

The global decision depends on the threshold values and the weight vector. The reader should note that the weights are assigned to the quantization bins, not to the reporting nodes.

Having a weight vector enables us to tune the system performance and the system behavior. Clearly, $\vec{w}$ partitions $\mathcal{B}$ (the set of all possible $\vec{B}$ vectors as depicted in Fig. 3.4) into two disjoint sets as $\mathcal{B}_0$ and $\mathcal{B}_1$, where the final decision is H_0 and H_1, respectively. $\delta_{\vec{w}}(.)$ evaluates to 0 in $\mathcal{B}_0$, and its corresponding terms in the probability evaluations become 0. Both $\mathbf{P_D}$ and $\mathbf{P_F}$ are evaluated by summing the probabilities of observing the cases in $\mathcal{B}_1$ since $\mathbf{P_D}$ and $\mathbf{P_F}$ are the probabilities of deciding H_1 as the final decision conditioned on having primary communication and no primary communication, respectively.

$$\mathbf{P^{ch}_{F,CSDV}} = \sum_{\vec{B} \in \mathcal{B}_1} (\mathbf{P}^{ch}_{H_0}(\vec{B})) \tag{3.17}$$

$$\mathbf{P^{ch}_{D,CSDV}} = \sum_{\vec{B} \in \mathcal{B}_1} (\mathbf{P}^{ch}_{H_1}(\vec{B})) \tag{3.18}$$

If we focus on the Rayleigh channel since it includes multipath effects as mentioned in Sect. 3.2.2, and using CSDV method as the fusion strategy leads to the following probabilities of cooperative detection and false alarm under a Rayleigh channel:

$$\mathbf{P^R_{F,CSDV}} = \sum_{\vec{B} \in \mathcal{B}_1} \mathbf{P}\{\text{Collected local decisions at fusion center} = \vec{B} | H_0\}$$

$$= \sum_{\vec{B} \in \mathcal{B}} \delta_{\vec{w}}(\vec{B}) \prod_{k=1}^{n} \binom{N - \sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}^R_{H_0}(B_k))^{N_{B_k}} \tag{3.19}$$

$$\mathbf{P^R_{D,CSDV}} = \sum_{\vec{B} \in \mathcal{B}_1} \mathbf{P}\{\text{Collected local decisions at fusion center} = \vec{B} | H_1\}$$

$$= \sum_{\vec{B} \in \mathcal{B}} \delta_{\vec{w}}(\vec{B}) \prod_{k=1}^{n} \binom{N - \sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}^R_{H_1}(B_k))^{N_{B_k}} \tag{3.20}$$

where N is the number of cooperative users, N_{B_k} is the number of users having observation in bin B_k, and n is the number of quantization bins.

3.5   Effects of Imperfect Reporting Channel

Let's also consider the cooperative sensing methods under an imperfect reporting channel. The effects of the imperfect reporting channel to the sensing performance are analyzed for hard decision logic functions in [38]. In [2], detection and false alarm probabilities are derived with consideration of errors in the reporting channel due to fading. Only the hard decision logic M-out-of-N is studied in the paper. In [8] Bit Error Probability (BEP) wall is introduced and performance analysis for M-out-of-N is done. The authors analyze the SNR loss due to BEP without considering the overall detection probabilities.
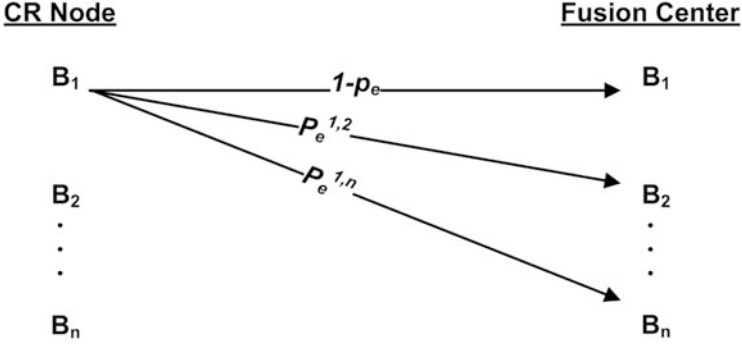
Fig. 3.5 Report channel model of n-bin quantization system

When the reporting channel is imperfect, errors may occur on the reported local decisions. In case of 1-bit quantization, L_i can only be 0 or 1 (corresponding to B_1 or B_2). Thus, it can be modeled as a *Binary Asymmetric Channel (BAC)* with total error probability p_e.

In the 1-bit quantization (hard decision) case, under H_1, the fusion center can receive 1 from a CR node in two situations: In the first case, CR node decides H_1 (B_2) locally and transmits it to the fusion center without any errors, which occurs with probability $\mathbf{P}_{H_1}^{ch}(B_2)(1 - p_e)$. The second situation occurs when CR node decides H_0 (B_1) locally and transmits it to the fusion center with error, which occurs with probability $\mathbf{P}_{H_1}^{ch}(B_1)p_e$. Therefore, the detection probability for a single node with an imperfect reporting channel becomes

$$\mathbf{P}_{H_1}^{ch,e}(B_2) = \mathbf{P}_{H_1}^{ch}(B_2)(1 - p_e) + \mathbf{P}_{H_1}^{ch}(B_1)p_e. \tag{3.21}$$

Using similar reasoning and formulation, the local false alarm probability with an imperfect reporting channel is given as

$$\mathbf{P}_{H_0}^{ch,e}(B_2) = \mathbf{P}_{H_0}^{ch}(B_2)(1 - p_e) + \mathbf{P}_{H_0}^{ch}(B_1)p_e. \tag{3.22}$$

In case of more quantization bins, the problem gets complicated. An example system model for n-bin quantization is depicted in Fig. 3.5. Only the arrows originating from B_1 are shown to make the figure more readable. The sum of error cases is denoted by p_e while $p_e^{i,j}$ represents the probability of receiving B_j at the fusion center when B_i is observed by a CR node and sent over the reporting channel. Here, the symbol errors occur and the symbol error rate for each symbol pair may be different according to the coding used.

Generalizing the formulation in Eqs. 3.21 and 3.22, the probability of receiving B_i at the fusion center through the imperfect reporting channel can be denoted as $\mathbf{P}_{H_1}^{ch,e}(B_i)$ and $\mathbf{P}_{H_0}^{ch,e}(B_i)$ under H_1 and H_0, respectively, and evaluate

$$\mathbf{P}_{H_1}^{ch,e}(B_i) = \mathbf{P}_{H_1}^{ch}(B_i)(1 - p_e) + \sum_{k \neq i} \mathbf{P}_{H_1}^{ch}(B_k) p_e^{k,i}, \qquad (3.23)$$

$$\mathbf{P}_{H_0}^{ch,e}(B_i) = \mathbf{P}_{H_0}^{ch}(B_i)(1 - p_e) + \sum_{k \neq i} \mathbf{P}_{H_0}^{ch}(B_k) p_e^{k,i}. \qquad (3.24)$$

After deriving these equations, one can plug $\mathbf{P}_{H_0}^{ch,e}(B_i)$ and $\mathbf{P}_{H_1}^{ch,e}(B_i)$ into $\mathbf{P_F}$ and $\mathbf{P_D}$ formulation in Eqs. 3.17 and 3.18, respectively. $\mathbf{P_F}$ and $\mathbf{P_D}$ for the CSDV under a Rayleigh and imperfect reporting channel can be evaluated as

$$\mathbf{P}_{\mathbf{F,CSDV}}^{\mathbf{R}} = \sum_{\vec{B} \in \mathcal{B}} \delta_{\vec{w}}(\vec{B}) \prod_{k=1}^{n} \binom{N - \sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}_{H_0}^{R,e}(B_k))^{N_{B_k}} \qquad (3.25)$$

$$\mathbf{P}_{\mathbf{D,CSDV}}^{\mathbf{R}} = \sum_{\vec{B} \in \mathcal{B}} \delta_{\vec{w}}(\vec{B}) \prod_{k=1}^{n} \binom{N - \sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}_{H_1}^{R,e}(B_k))^{N_{B_k}} \qquad (3.26)$$

where $\mathcal{B}$ is the set of all possible $\vec{B}$s, N is the number of cooperative users, N_{B_k} is the number of reports in bin B_k from the fusion center's perspective (with errors), $\mathbf{P}_{H_i}^{R,e}(B_k)$ is the probability of having a report in bin B_k under H_i in a Rayleigh channel, and n is the number of quantization bins.

3.6 Optimizing Detector That Uses Quantization

Recall that the main goal of spectrum sensing is to decide between two hypotheses that the channel state is empty (H_0) or it is actively used (H_1). If the conditional distributions of O_i are known, Neyman-Pearson optimality that maximizes $\mathbf{P}_{\mathbf{D,CSDV}}^{\mathbf{ch}}$ subject to the $\mathbf{P}_{\mathbf{F,CSDV}}^{\mathbf{ch}} < \alpha$ constraint can be used.

When we use quantization and a functional global decision logic, we need to optimize λ_1, Δ, and $\vec{w}$. We portray two problem definitions. The first problem aims to optimize the thresholds of the quantizer for a given $\vec{w}$ and α. The second problem aims to optimize the $\vec{w}$, however $\mathbf{P}_{\mathbf{D,CSDV}}^{\mathbf{ch}}$ is not differentiable with respect to $\vec{w}$.

3.6.1 Threshold Optimization

The optimization problem for optimizing λ_1 and Δ under a Rayleigh channel when there are no false reports and the perfect reporting channel is formulated in Eq. 3.27
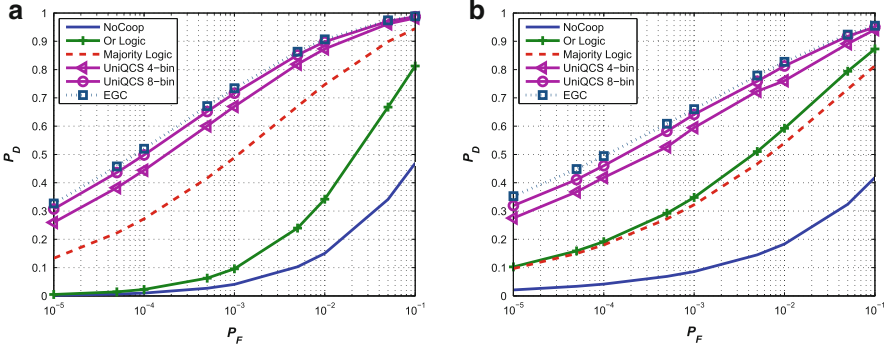
Fig. 3.6 ROC curves of different global decision logic functions (SNR = 5 dB, TW = 5, $N = 10$). (**a**) AWGN channel. (**b**) Rayleigh channel

when $\overrightarrow{w}$ is given. When the reporting channel is imperfect we can obtain similar optimization problem formulations by using Eqs. 3.25 and 3.26, respectively.

$$UniQCS1: \underset{\Delta,\lambda_1}{maximize} \sum_{\overrightarrow{B}\in\mathcal{B}} \delta_{\overrightarrow{w}}(\overrightarrow{B}) \prod_{k=1}^{n} \binom{N\text{-}\sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}_{H_1}^{R}(B_k))^{N_{B_k}}$$

$$subject\ to\ \sum_{\overrightarrow{B}\in\mathcal{B}} \delta_{\overrightarrow{w}}(\overrightarrow{B}) \prod_{k=1}^{n} \binom{N\text{-}\sum_{j=1}^{k-1} N_{B_j}}{N_{B_k}} (\mathbf{P}_{H_0}^{R}(B_k))^{N_{B_k}} \le \alpha \quad (3.27)$$

3.6.2 Improving Weights and Threshold Optimization

We introduce another optimization problem for optimizing $\overrightarrow{w}$ similar to UniQCS1.

$$UniQCS2: \underset{\overrightarrow{w}}{maximize}\ \mathbf{P_D}\ subject\ to\ \mathbf{P_F} \le \alpha \quad (3.28)$$

Different $\overrightarrow{w}$ vectors change $\mathcal{B}_1$, hence $\delta_{\overrightarrow{w}}(\overrightarrow{B})$ values. This change further affects $\mathbf{P_D}$ values evaluated optimally by UniQCS1.

3.6.3 Results

In Fig. 3.6a, b, the performance of different global decision logic functions is analyzed by using ROC curves under AWGN and Rayleigh channels, respectively.

Fig. 3.7 P_e versus P_D curves of different global decision logic functions including imperfect reporting channel ($P_F = 0.01$, SNR = 5 dB, TW = 5, $N = 10$)
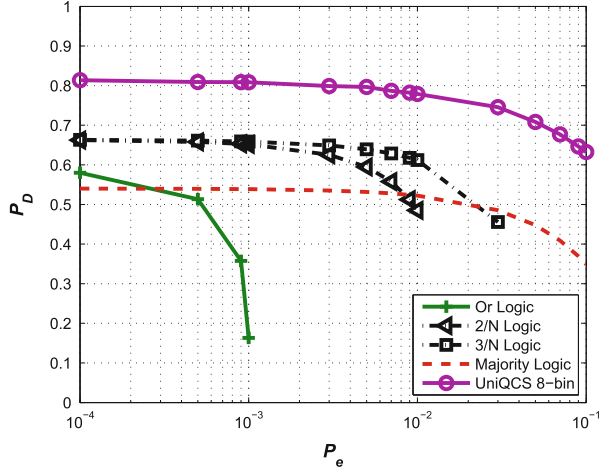
In the figures, P_D values corresponding to the given P_F constraint are plotted. UniQCS that utilizes uniform quantizer for cooperative sensing performs close to EGC with 8-bin quantization and far better than the hard decision logic functions. For $P_F = 10^{-2}$, UniQCS 8-bin achieves 0.9 and 0.81 P_D under AWGN and Rayleigh channels, respectively. For the same P_F, Majority logic achieves 0.75 and 0.54, respectively.

Effect of P_e on P_D is analyzed in Fig. 3.7. Increasing P_e decreases P_D, if $P_F = 0.01$ is achievable. UniQCS and Majority logic are robust against an imperfect reporting channel. However, the Majority logic performance is much lower than that of UniQCS.

3.7 Cooperative Sensing Scheduling

Sensing scheduling is the selection and assignment of SUs to the channels to be sensed together with the determination of operational parameters for the sensing operation (e.g. sensing duration, sensing threshold, etc.). Cooperative sensing scheduling (CSS) is the case where sensing scheduling is performed in a way that leverages sensing performance by selecting the best set of cooperating SUs for each channel.

In the previous section, methods for spectrum sensing in a collaborative manner have been discussed. Such methods boost the probability of detection and lower the probability of false alarm by utilizing the diversity of the measurements from multiple SUs. However, the large number of channels to be sensed multiplied by the minimum number of sensing SUs for each channel to allow for collaboration proves that plenty of SUs should perform sensing all over the service area at anytime. Requiring almost all SUs to perform sensing continuously results in enormous

energy expenditure and the mobile SU devices running out of battery. Therefore, the channels to be sensed as well as the set of SUs that should sense each channel must be selected carefully. CSS targets to do this selection effectively in order to improve the performance of the cooperative spectrum sensing

3.7.1 Introduction

In this section besides improving detection probability by intelligent assignment of SUs to sense selected channels, we also consider the energy efficiency problem from the mobile user's perspective in an infrastructure based cognitive radio network (CRN). In a multi-channel CRN, the cognitive base station (CBS) decides on the set of SUs that should sense the individual channels in order to ensure high detection probability and low false alarm probability. We take into account the individual SNR values of SUs, as a high SNR value implies less energy consumption and better performance. Most of the previous work in the literature consider detection accuracy as the main performance criterion whereas energy efficiency with satisfactory primary user protection should also be an important goal.

The authors of [25,50,63] focus on the energy efficiency of CSS and formulate it as a combinatorial optimization problem. As these works do not differentiate SUs in terms of SNR, the problem boils down to simply deciding on the number of SUs to sense a channel. Sun et al. [50] assume identical SNR and sensing duration for each SU. Similarly, Zhang et al. [63] decide on the number of SUs and sensing duration. Hao et al. [25] propose a method where each SU decides the set of channels to be sensed individually. Unlike these works, a heterogeneous environment in terms of SNR of the primary user's signal is considered in this section as it differs among SUs due to geographical conditions, multi-path, etc. An optimization model that decides the sensing set of SUs for each primary channel together with the sensing duration for each SU/channel pair is proposed [16].

3.7.2 System Model

The system operates in a frame based manner. At the beginning of each frame, there is a quiet sensing period of length T^s during which all SUs cease communication and sense the channels assigned to them by the CBS. During this period, an SU may sense more than one channel by switching from one channel to another. However, channel switching also requires time and brings an energy overhead. Thus, the order of the channels is also important. An SU that finishes its sensing task switches to a low-power idling mode in order to save energy. After this sensing period, there is a reporting period where the SUs send their local sensing hard decisions (0 or 1) to CBS for fusion. CBS performs fusion operation using OR rule, and decides on the transmission schedule based on the gathered data, which is broadcasted to SUs.
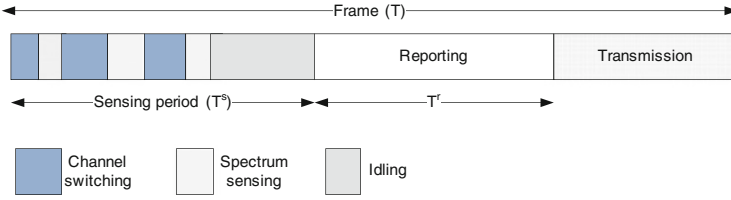
Fig. 3.8 Frame organization

The final part of the frame is reserved for transmission, during which SUs send their packets following the schedule. The organization of the frame can be seen in Fig. 3.8. The main focus is on the part of the frame related to sensing (sensing period and reporting period) [16].

Let M, N, and $\gamma_{m,n}$ denote the number of primary channels, number of SUs, and SNR value of SU_n over channel m, respectively. It is assumed that $N \gg M$. In order to ensure diversity and cooperation, the requirement is that each channel should be sensed by at least δ^{min} SUs. All M channels should be sensed with sufficient accuracy while consuming as little energy as possible.

3.7.2.1 Energy Consumption Model

For a CRN, energy consumed for the sensing task can be factored into three components:

- Sensing Energy (E^s): This is the energy that is spent by an SU for sensing a given channel. If we let P^s and $\tau_{m,n}$ be the sensing power and sensing duration of SU_n for channel m respectively, then E^s can be calculated as $\sum_{m=1}^{M} \sum_{n=1}^{N} P^s \tau_{m,n}$.
- Channel Switching Energy (E^{cs}): This is the energy consumed when an SU switches from one channel to another during the sensing period. The power required for this operation is defined as channel switching power (P^{cs}) and the related energy consumption is given as $P^{cs} T^{cs}$ where T^{cs} is the total time required for completing the channel switch. T^{cs} is given by $t^{cs} |f - f'|$ where t^{cs} is the time required to switch to the adjacent channel and $|f - f'|$ is the absolute value of separation between the two frequencies [23]. Let f_n^0 denote the frequency of the channel to which the antenna of the SU_n is tuned at the beginning of the quiet period, and C_n be the ordered set of frequencies that are going to be sensed by SU_n, i.e., $C_n = \{f_n^1, f_n^2, \ldots\}$. Then $E^{cs} = P^{cs} t^{cs} \sum_{n=1}^{N} (|f_n^0 - f_n^1| + \sum_{k=1}^{|C_n|-1} |f_n^k - f_n^{k+1}|)$ where $|C_n|$ is the cardinality of C_n.
- Reporting Energy (E^r): This is the energy that is consumed by an SU for sending its local sensing decisions to CBS for fusion. It is assumed that regardless of the number of channels sensed, an SU only sends a single packet if it senses anything.

Moreover, the reporting period is long enough such that all SUs can send their reporting packets. If we let E_n^{tx} be the energy required by SU_n to send a single packet to CBS, and $\mathcal{S}$ be the set of reporting SUs, E^r is given by $\sum_{n \in \mathcal{S}} E_n^{tx}$.

3.7.3 Problem Formulation

Let's assume that all SUs have the same false alarm probability for all channels, denoted by (P^F). The probability of detection of SU_n for channel m is given by

$$P_{m,n}^D = \mathcal{Q}\left(\frac{\mathcal{Q}^{-1}(P^F) - \sqrt{\tau_{m,n} f_s} \gamma_{m,n}}{\sqrt{2\gamma_{m,n} + 1}}\right) \tag{3.29}$$

where f_s is the sampling frequency and $\mathcal{Q}$ is the complementary cumulative distribution of a standard Gaussian [33]. The cooperative detection probability for channel m using OR rule, Q_m^D, is given as

$$Q_m^D = 1 - \prod_{n \in \mathcal{S}_m} (1 - P_{m,n}^D) \tag{3.30}$$

where $\mathcal{S}_m$ is the set of SUs sensing channel m. In a previous work [17], the authors showed that the cooperative detection probability for channel m is a concave function of $\tau_{m,n}$ if all the individual detection probabilities of SUs participating in the sensing of channel m are greater than 0.5. Let $\tau_{m,n}^{min}$ denote the minimum time required to achieve a $P_{m,n}^D$ value of 0.5, that is given by

$$\tau_{m,n}^{min} = \left(\frac{\mathcal{Q}^{-1}(P^F)}{\gamma_{m,n}\sqrt{f_s}}\right)^2. \tag{3.31}$$

In addition, to ensure a cooperative false alarm threshold of $_{th}Q^F$, a channel should be sensed by at most $\lfloor \frac{\log(1 - _{th}Q^F)}{\log(1 - P^F)} \rfloor$ number of SUs.

The sensing sequence (the set of sensed channels and their order) of an SU is modeled by using network flows. Let ϕ denote a virtual terminal channel that indicates the end of sensing, and f_m be the frequency of channel m. Figure 3.9 illustrates the network flow representation for the sensing actions for an SU. SU_n's antenna is tuned to f_n^0 at the beginning of a frame. Moreover, each SU switches to virtual channel ϕ when its sensing task is done. As there are M channels, an SU can perform at most $M + 1$ switches. However, it can also sense less channels resulting in an SU to have an outgoing arrow directly after kth step of sensing.

The optimization model that minimizes energy consumption related to sensing is given subsequently. Let $\tau_{m,n}$ be a non-negative continuous variable (i.e., $\tau_{m,n} \geq 0$)
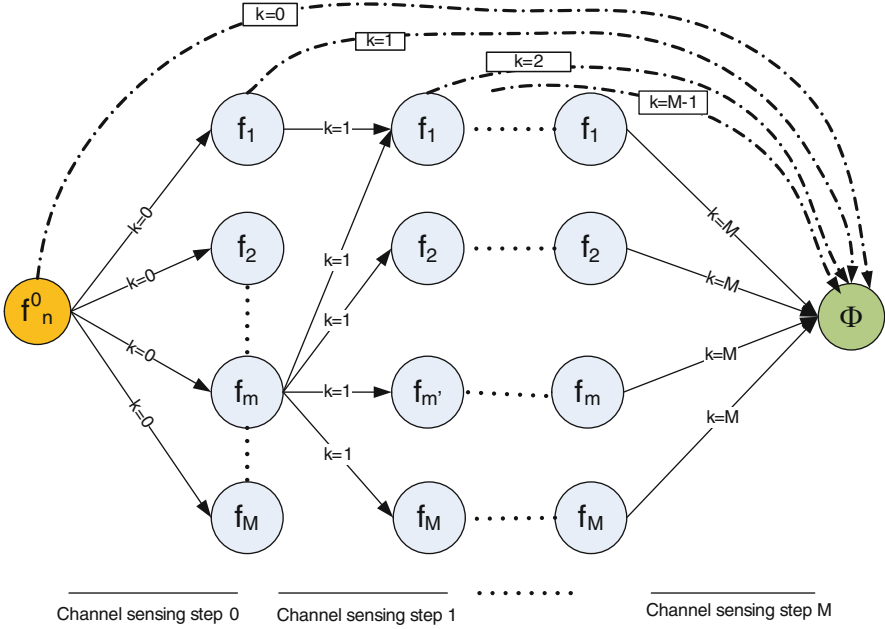
Fig. 3.9 Channel sensing sequence

denoting the time SU_n spends for sensing channel m, $x_{m,m',n}^k$ be a binary variable (i.e., $x_{m',m,n}^k \in (0,1)$) with value 1 if SU_n switches from f_m to f_m' at step k, and y_n be a binary variable (i.e., $y_n \in (0,1)$) with value 1 if SU_n transmits its sensing outcomes to the base station. Then the optimization model can be written as:

P1: $\min \ w = \sum_{m=1}^{M} \sum_{n=1}^{N} P^s \tau_{m,n} + \sum_{n=1}^{N} E_n^{tx} y_n$

$$+ P^{cs} t^{cs} \sum_{n=1}^{N} (\sum_{m=1}^{M} |f_n^0 - f_m| x_{f_n^0,m,n}^0 \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{k=1}^{M-1} |f_m - f_{m'}| x_{m,m',n}^k)$$

$$(3.32)$$

Flow related constraints:

$$x_{f_n^0,\phi,n}^0 + \sum_{m=1}^{M} \sum_{k=1}^{M} x_{m,\phi,n}^k = 1, \ \forall n \tag{3.33}$$

$$x^0_{f^0_n,\phi,n} + \sum_{m=1}^{M} x^0_{f^0_n,m,n} = 1, \ \forall n \tag{3.34}$$

$$x^0_{f^0_n,m,n} - (x^1_{m,\phi,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} x^1_{m,m',n}) = 0, \ \forall m, \forall n \tag{3.35}$$

$$\sum_{\substack{m'=1 \\ m'\neq m}}^{M} x^{M-1}_{m',m,n} - x^M_{m,\phi,n} = 0, \ \forall m, \forall n \tag{3.36}$$

$$\sum_{\substack{m'=1 \\ m'\neq m}}^{M} x^k_{m',m,n} - \sum_{\substack{m'=1 \\ m'\neq m}}^{M} x^{k+1}_{m,m',n} = 0, \ \forall m, \forall n, k = 1, \ldots, M-2 \tag{3.37}$$

$$x^0_{f^0_n,m,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{k=1}^{M-1} x^k_{m',m,n} \leq 1, \ \forall m, \forall n \tag{3.38}$$

Sensing time related constraints:

$$\tau_{m,n} - \tau^{min}_{m,n}(x^0_{f^0_n,m,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{k=1}^{M-1} x^k_{m',m,n}) \geq 0, \ \forall m, \forall n \tag{3.39}$$

$$t^{cs} \sum_{m=1}^{M} (|f^0_n - f_m| x^0_{f^0_n,m,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{k=1}^{M-1} |f_m - f_{m'}| x^k_{m,m',n}) + \sum_{m=1}^{M} \tau_{m,n} \leq T^s y_n, \ \forall n$$
$$\tag{3.40}$$

Sensing quality related constraints:

$$\sum_{n=1}^{N} (x^0_{f^0_n,m,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{k=1}^{M-1} x^k_{m',m,n}) \geq \delta^{min}, \ \forall m \tag{3.41}$$

$$\sum_{n=1}^{N} (x^0_{f^0_n,m,n} + \sum_{\substack{m'=1 \\ m'\neq m}}^{M} \sum_{k=1}^{M-1} x^k_{m',m,n}) \leq \lfloor \frac{\log(1 - {}_{th}Q^F)}{\log(1 - P^F)} \rfloor, \ \forall m \tag{3.42}$$

$${}_{th}Q^D - Q^D_m \leq 0, \qquad \forall m \tag{3.43}$$

where Q_m^D is calculated as follows:

$$Q_m^D = 1 - \prod_{n=1}^{N} \left(1 - P_{m,n}^D (x_{f_n^0,m,n}^0 + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sum_{k=1}^{M-1} x_{m',m,n}^k) \right).$$

The objective in (3.32) minimizes the total energy expenditure due to sensing for a frame. The first part of the third term is the energy consumption due to the initial channel switch from f_n^0 whereas the second part is for the succeeding channel switches. Constraint (3.33) indicates that the sensing of all SUs should end at some step k. In other words, the incoming flow to node ϕ should be 1. Constraint (3.34) makes sure that sensing sequence of each SU begins with step 0 (The outgoing flow from node f_n^0 should be 1). Constraints (3.35)–(3.37) are flow conservation equations for step 0, step M, and intermediate steps, respectively. Constraint (3.38) states that an SU can sense a given channel at most once for a frame. Constraint (3.39) enforces $\tau_{m,n}$ to be greater than or equal to $\tau_{m,n}^{min}$ if SU_n switches to channel m, which is required the concavity condition for Q_m^D to hold. Constraint (3.40) expresses the fact that the total time spent by each SU should be less than the length of the quiet sensing period. Constraint (3.41) forces each channel to be sensed by at least δ^{min} SUs. On the other hand, Constraint (3.42) makes sure that the cooperative false alarm probability be smaller than the respective threshold for all channels. Constraint (3.43) is the cooperative detection probability constraint [16].

3.7.3.1 Outer Linearization (OL)

The given model is a mixed integer non-linear problem with a linear objective. However, it is convex once binary variables are fixed as Q_m^D is concave in terms of $\tau_{m,n}$. Thus, outer linearization algorithm is employed, which first ignores all the non-linear constraints, then iteratively linearizes the violated ones by using the gradient until all of them are satisfied [11]. The steps of the algorithm are as follows:

Step 1: Let h denote the iteration number, and set $h = 1$. Ignore the non-linear constraints of the original problem, solve the relaxed problem (**P2**) to obtain a solution $\tau_{m,n}^h, x_{m',m,n}^{k,h}, y_n^h$.

Step 2: Find the most violated constraint among the M previously ignored constraints, say constraint θ with the current solution, $\tau_{m,n}^h, x_{m',m,n}^{k,h}, y_n^h$. Let us call the maximum violation v_θ, and the corresponding constraint g_θ. If $v_\theta < \epsilon$, then the current solution is optimal with ϵ feasibility tolerance. Otherwise, proceed with the next step.

Step 3: Linearize g_θ by adding the following constraint to **P2**:
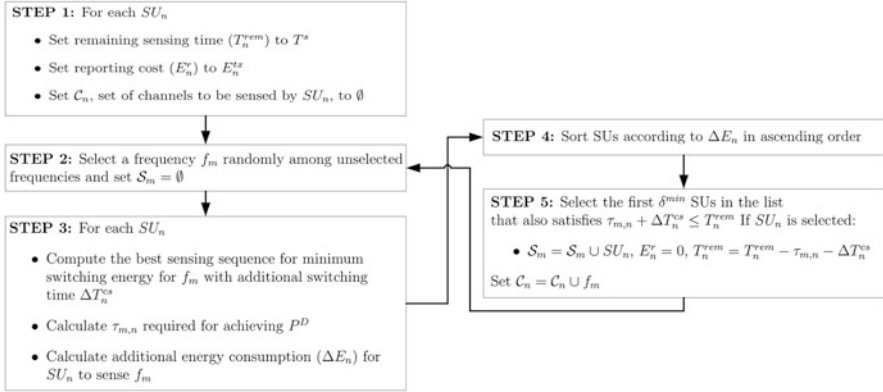
Fig. 3.10 EASE algorithm flow diagram

$$\nabla g_\theta(\ldots x^{k,h}_{m',\theta,n}, \ldots \tau^h_{\theta,n}, \ldots)^T \begin{pmatrix} \vdots \\ x^{k}_{m',\theta,n} - x^{k,h}_{m',\theta,n} \\ \vdots \\ \tau_{\theta,n} - \tau^h_{\theta,n} \\ \vdots \end{pmatrix} + v_\theta \leq 0$$

where $\nabla g_\theta(\ldots x^{k,h}_{m',\theta,n}, \ldots \tau^h_{\theta,n}, \ldots)$ is the gradient vector of g_θ evaluated at the current solution.

Step 4: Set $h = h + 1$, solve the modified mixed integer linear problem to obtain a new solution, and proceed with Step 2.

3.7.3.2 A Low Complexity Heuristic Algorithm: Energy Aware Sensing Scheduling (EASE)

EASE is greedy construction heuristic that assigns δ^{min} SUs to sense a channel such that all assigned SUs have identical probability of detection, P^D, that is calculated as $1 - (1 - {}_{th}Q^D)^{1/\delta^{min}}$ for OR rule in order to satisfy cooperative detection probability constraint. The motivation behind the heuristic that is shown in Fig. 3.10, is to select the set of SUs that will consume the least additional energy for sensing each channel. At each iteration (i.e., assignment for f_m), SUs are sorted in ascending order based on their additional energy requirement to sense the channel (ΔE_n) which includes switching, sensing and reporting costs. EASE assigns SUs to channels sequentially and updates the channel sensing sequence

Table 3.1 Model parameters

δ^{min}	3	E_n^{tx}	1 mJ, $\forall n$
T^s	20 ms	μ^{SNR}	Between -5 and 0 dB
P^F	0.01	t^{cs}	Between 0.5 and 1.5 ms per 100 kHz
f_s	1 kHz	$_{th}Q^D$	0.9
P^s	1,000 mW	$_{th}Q^F$	0.1
P^{cs}	1,000 mW	ϵ	10^{-4}
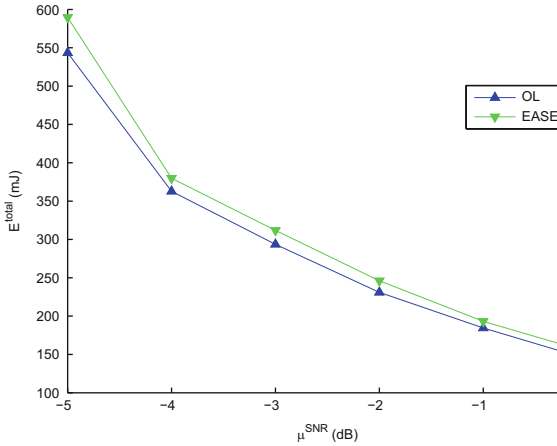
Fig. 3.11 Energy vs μ^{SNR} with t^{cs}=1 ms/100 kHz

to obtain the minimum switching energy. Moreover, an SU that is not currently assigned a channel also incurs a reporting cost. Among the sorted SUs, the first δ^{min} ones with enough remaining time are assigned to that channel. The complexity of the heuristic is $O(MNlogN)$.

3.7.4 Performance Evaluation

The performance of both methods is analyzed in a network with 20 contiguous channels with 100 kHz bandwidth each and 100 SUs. f_n^0 are assigned randomly. Furthermore, it is assumed that $\gamma_{m,n}$ values are exponentially distributed with mean μ^{SNR}. For both f_n^0 and $\gamma_{m,n}$, the same randomly assigned values are used for consistency. The other parameters are given in Table 3.1. The results given below are for a single frame. Hence, cumulative energy savings in the long run will be much higher. As the results of EASE depend on the order of the channels for assignment, we run EASE with 20 random orderings and give the results for the ordering with the minimum total energy consumption.

The effect of μ^{SNR} on total energy consumption is shown in Fig. 3.11. Increasing μ^{SNR} value results in less energy expenditure since less time is required to achieve

Fig. 3.12 Energy consumption profiles with $t^{cs} = 1\,\text{ms}/100\,\text{kHz}$: $\mu^{SNR} = -5\,\text{dB}$
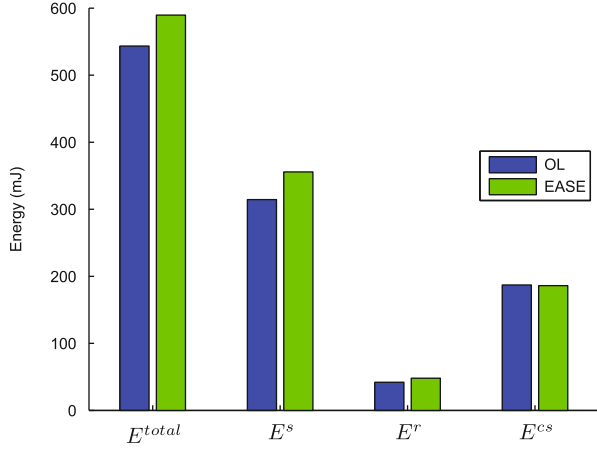
Fig. 3.13 Energy consumption profiles with $t^{cs} = 1\,\text{ms}/100\,\text{kHz}$: $\mu^{SNR} = 0\,\text{dB}$
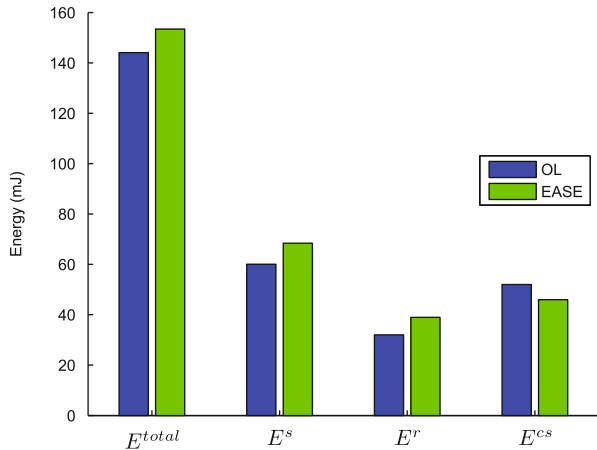
a given detection probability. However, the gains for improving μ^{SNR} show a decreasing pattern. Moreover, EASE performs very close to OL, always within 10 %.

Total energy consumption and its components are shown in Figs. 3.12 and 3.13 for low and high μ^{SNR} values. In both cases, E^r remains almost the same. For low μ^{SNR}, sensing energy is the dominant factor, whereas for high μ^{SNR}, all components are close to each other. Furthermore, although E^{cs} is not directly related to μ^{SNR}, increasing μ^{SNR} also decreases E^{cs}, as more SUs become candidates for sensing a particular channel within sensing period, which in turn provides better sensing sequences.

Finally, E^{total} and its components for low and high t^{cs} values are shown in Figs. 3.14 and 3.15. Increasing t^{cs} increases both E^{cs} and E^s. E^{cs} is linearly related to t^{cs}, hence, the increase is expected. On the other hand, a higher value of t^{cs} leaves less time for sensing, which decreases the number of candidate channels for sensing

Fig. 3.14 Energy
consumption profiles with
$\mu^{SNR} = -3$ dB:
$t^{cs} = 0.5$ ms/100 kHz
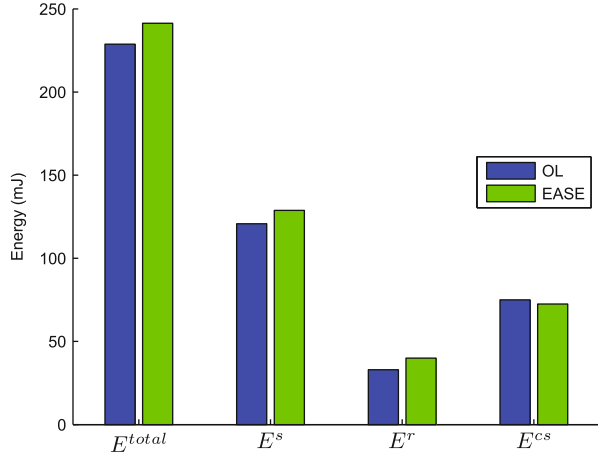
Fig. 3.15 Energy
consumption profiles with
$\mu^{SNR} = -3$ dB:
$t^{cs} = 1.5$ ms/100 kHz
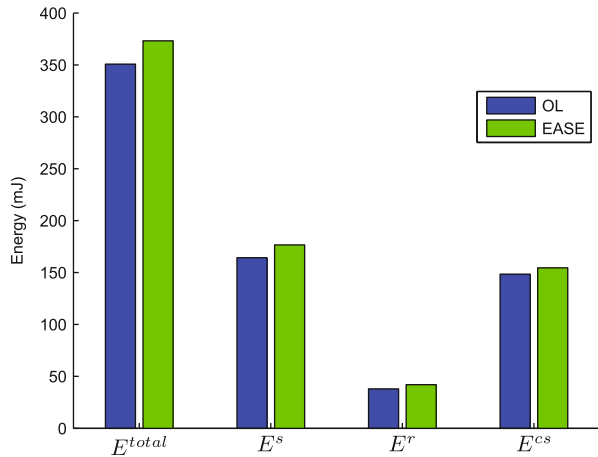
a given channel. Moreover, with increasing t^{cs}, some SUs may refrain to switching
to a channel. Thus, E^s also increases with t^{cs} [16].

3.8 Utilizing Sensing Results for REM Construction

Cognitive radio research community is lately inclined on utilizing white space
database lookups to decide on the availability of radio spectrum for different
geographies. This approach allow the use of historical spectrum usage information
in a specific area as well as protecting registered technologies, such as wireless
microphone systems. Although the use of white space databases proves to be

practical, the construction of such databases for spectrum bands other than that of the wireless microphones remains as an open research problem.

This is where REM comes into play. REM, without requiring any a priori registration for spectrum usage, can locate PUs for any spectrum band and estimate the region where SUs should abstain from utilizing that band. Thus, REM constructs a map of multiple layers where each layer represents a temperature map of a specific spectrum band of interest. This multi-layer REM enables the realization of the cognitive engine for cognitive radio. For each location, it is now possible to check individual layers of REM and select the most appropriate band for the SU.

The cooperative spectrum sensing and sensing scheduling methods discussed earlier in this chapter can be utilized to construct the layers of REM. For each layer, i.e., for each spectrum band, the sensing scheduling method selects the set of SUs that should cooperate to sense that spectrum band. The observations of those SU devices is then utilized by the cooperative sensing method to construct the temperature map for that band. Initiated by the FCC in the USA, the use of TVWS for dynamic spectrum access has attracted worldwide interest such as the UK, Brazil, Japan, India, Singapore, and China. DSO is almost completed in Japan except for the three prefectures devastated by the March 11, 2011 earthquake, in Canada, and some parts of Europe. It is completely finished in most of the Europe by 2012. South Korea, Australia, China, and Brazil are expected to switch off not later than 2015 [35]. Turkey launched trial transmissions in 2006 and originally planned to gradually do the switch by 2014 [53].

As the first step of opportunistic spectrum access, a TVWS device must establish the available frequencies at its location just before communications and must ensure the PUs are not harmed. Due to the concerns mainly raised by the broadcasting industry on the proper protection of the incumbent TV services, regulatory bodies have brought strict rules on the white space devices. The rules demanded by the regulators such as the detection of a primary signal at -114 dBm by FCC are over-conservative. Such regulations reduce the available white spaces drastically (approximately by a factor of 3) [26]. Instead, accessing a centralized database, which keeps track of the available spectrum based on geographical coordinates or the properties of registered PU transmitters, has been accepted more promising for mitigating the strict sensing challenges while meeting the regulatory obligations [36].

The Geolocation database stores TV base station locations and the related parameters (static data) and constructs a real-time view of the spectrum occupancy at the TV bands at each location [36], which is a quasi-static information. Currently, the database based TVWS operation is being pursued by regulators in the USA and the UK. In November 2009, FCC opened a call for database administrators and approved ten companies (Google, Spectrum Bridge, Telcordia, Microsoft, etc.) as administrators in 2010. A step further from the geolocation database lies the radio environment map. The REM was first defined as an abstraction of a real-world environment storing multi-domain information dynamically [66]. The REM can act
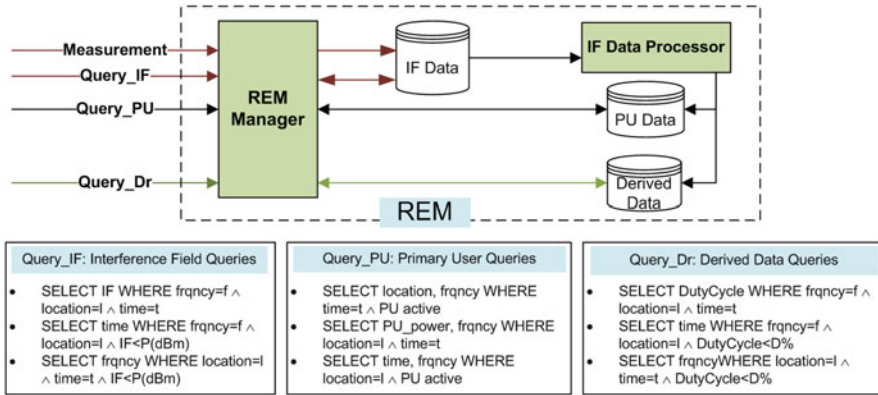
Fig. 3.16 REM data flow and example queries

as an enabler for cognition in radios; it can store PU activity statistics as well as RF environment information such as propagation characteristics even if it is dynamic information.

The REM was first defined as an abstraction of a real-world environment storing multi-domain information [66]. However, it can also be considered more generally as an intelligent network entity that can further process the gathered information, inspect the spatio-temporal characteristics, and derive a map of the RF environment [48]. The REM is a promising concept for efficient CRN operation without extensive burden on CRs as it can be considered as the *cognitive engine* located at the network. The REM introduces environment awareness that would be harder to acquire by individual CR capabilities via extensive spectrum analysis. Hence, the REM can also be seen as the network support turning simple nodes into intelligent CRs. The design of the REM relies on the following related processes: data gathering and representation, data processing/fusion, and data retrieval/query as depicted in Fig. 3.16.

Essential functionality of a REM is the construction of a dynamic interference map for each frequency at each location of interest which is volatile data. Since it is impractical to have measurements at each location in the CR operation area, the REM fuses the available measurements to estimate the interference level at locations with no measurement data. The radio interference map interpolation is the most critical part of a REM construction method.

3.8.1 REM Architecture

The data stored in a REM depends on the measurements collected from CR nodes in the network and other Measurement Capable Devices (MCD). Each measurement

Table 3.2 The model of the data reported by the MCDs

Parameter	Size	Units
RSS	Vector	(dBm)
Frequency band	Vector	(MHz)
Time instance	Scalar	(s)
x-coordinate	Scalar	(Latitude)
y-coordinate	Scalar	(Longitude)
z-coordinate	Scalar	(Altitude)
MCD identifier	Label	(No dimension)
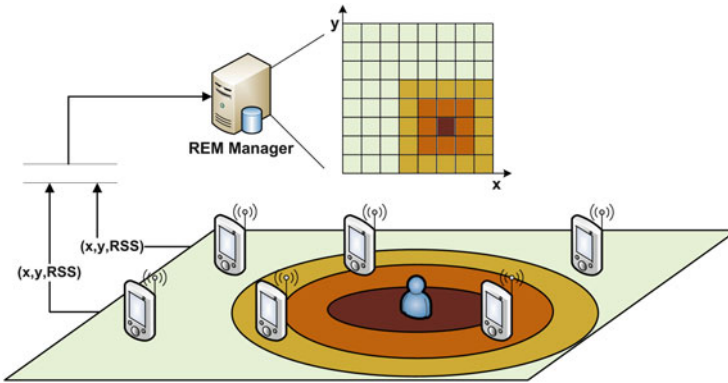
Fig. 3.17 System model of CR operation area

report also has the geolocation and time stamps. MCDs can be specific purpose dedicated sensor nodes, mobile phones, or similar devices.

The sensing data records should contain the following information:

- Received primary (and/or secondary) signal levels
- Location of the measurement (longitude, latitude, altitude)
- Time stamp
- Inspected frequency bands

The raw sensing data depicted in Table 3.2 (adapted from [13]) are reported to functional entity REM Manager that exploits the reported information and information from the geolocation DB to build a complete map by interpolating the geo-localized measurements. The REM Manager persists the raw data and the interpolated data to a database for responding queries easily and fast.

REM system model is depicted in Fig. 3.17. MCDs are placed in a grid-like layout in the operation area and a primary user is actively using a frequency band. These MCDs may be both devices deployed specifically by the cognitive radio operator or secondary user devices. Circular regions represent the different interference level contours. REM manager collects the $(x, y, RSS, \text{Time stamp})$ values from the measuring nodes and constructs the radio interference map via

Fig. 3.18 False alarm, missed detection, and correct detection zones
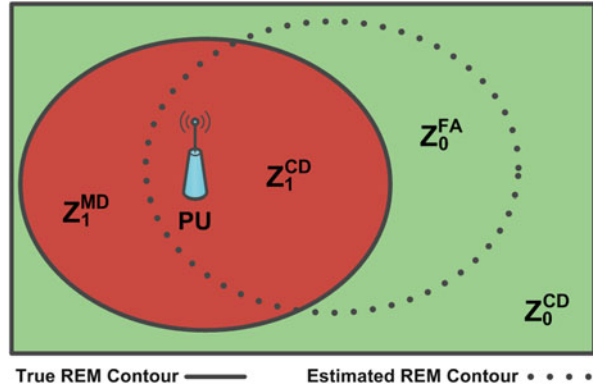
interpolating the RSS values at locations without any measurements. RSS values depend on the characteristics of the propagation environment.

3.8.2 REM Quality Metrics

How accurate the REM describes the real operation environment is measured by the REM quality metrics. These quality metrics can be evaluated by comparing the estimated REM with the true REM in terms of the correct detection zone, false alarm zone, and RMSE. Number of sensors, distribution of sensors in the field, capability of sensors, dynamics of the propagation environment, and accuracy of the propagation modeling are the key parameters affecting the performance of the REM.

The intersection of the zones of the true REM and the estimated REM defines the zones that are correctly and incorrectly determined. In Fig. 3.18, the true REM and the estimated REM contours are depicted. The solid line divides the area into two subareas in which the transmission by secondary users is banned and allowed, denoted by dark and light shaded areas. The subscripts of the zones depict the actual state with respect to the true REM. The areas where the true REM and the estimated REM contours both determine H_1 or H_0 zones correctly are called the Correct Detection Zone type-1 (Z_1^{CD}) and Correct Detection Zone type-0 (Z_0^{CD}), respectively. False Alarm Zone (Z_0^{FA}) is the area in which transmission is forbidden due to estimation errors though it would not harm the primary users. Finally, Missed Detection Zone (Z_1^{MD}) is the area where the transmission is not allowed but the estimated REM infers the opposite. *New metrics for the performance evaluation of the REM construction techniques are proposed in [60]. The regions according to true REM are normalized and the Z_1^{CD} and Z_0^{FA} ratios (CDZR$_1$ and FAZR) are determined, which present a better understanding of the performance of the REM construction algorithm;*

$$\text{FAZR} = \frac{A(Z_0^{FA})}{A(Z_0^{FA}) + A(Z_0^{CD})} \tag{3.44}$$

$$\text{CDZR}_1 = \frac{A(Z_1^{CD})}{A(Z_1^{CD}) + A(Z_1^{MD})} \tag{3.45}$$

where $A(Z)$ represents the area of zone Z. FAZR and CDZR$_1$ are analogous to $\mathbf{P_F}$ and $\mathbf{P_D}$ in the spectrum sensing scope. Similarly, we define MDZR as

$$\text{MDZR} = \frac{A(Z_1^{MD})}{A(Z_1^{CD}) + A(Z_1^{MD})}.$$

Analogy of $\mathbf{P_D} + \mathbf{P_M} = 1$ is reflected as CDZR$_1$ + MDZR = 1.

3.8.3   RSS Measurements in Fading Channels

RSS measurements in fading channels are mainly characterized by the fading type. We use log-normal shadowing and Suzuki process for slow fading and fast fading channels, respectively. Suzuki process includes log-normal shadowing and Rayleigh fading hence, we use shadowing/fading shorthand for the fast fading channel.

The RSS measurements dynamics for slow and fast fading channel conditions are known in the literature. Consequently, we model the mean RSS in dB scale for the MCDs and depending on the fading type. The fading can cause poor performance in a communication system because it can result in a loss of signal power without reducing the power of the noise. This signal loss can be over some or all of the signal bandwidth. Therefore, it is important to model the channel and mean RSS since it affects the system performance while estimating the location of the primary signal and constructing the REM.

3.8.3.1   RSS Measurements in Slow Fading Channel

Long term fading arises when the coherence time of the channel is large relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel can be considered roughly constant over the period of use. Slow fading can be caused by events such as shadowing, where a large obstruction such as a hill or large building obscures the main signal path between the transmitter and the receiver. The amplitude change caused by shadowing is often modeled using a log-normal distribution.

Under a log-normal channel, RSS measurements in dB obey normal distribution. Hence, the analysis in dB is more appropriate. The ideal RSS at the ith MCD, denoted by P_i^{rx}, is expressed as

$$P_i^{rx}[\text{dB}] = P^{tx}[\text{dB}] - P_{L_0} - 10\log_{10} d_{(x_i,y_i)}^{\alpha} + S_i \tag{3.46}$$

where P_{L_0} and α are path loss correction and path loss exponents, respectively. P^{tx} is the transmit power of the transmitter and $d_{(x_i,y_i)}$ is the distance between active primary user and the ith secondary user or MCD. S_i is a Gaussian random variable with mean zero and variance σ_s^2 for expressing the effect of log-normal shadowing.

RSS values at each MCD has severe disturbances caused by the shadowing effect, so we measure raw RSS values and evaluate sample mean to reduce the shadowing effect. Mean RSS is formulated as

$$\overline{P_i^{rx}}[\text{dB}] = \sum_{j=1}^{N_m} P_{i,j}^{rx}[\text{dB}]/N_m \tag{3.47}$$

where $P_{i,j}^{rx}$ is the jth measured RSS value at the ith MCD and N_m is the total number of samples for unit measurement. Since the mean of S_i is zero, mean RSS value at ith MCD has the mean $P^{tx}[\text{dB}] - P_{L_0} - 10\log_{10} d_{(x_i,y_i)}^{\alpha}$.

3.8.3.2 RSS Measurements in Fast Fading Channel

The short term fading occurs when the coherence time of the channel is small relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel fluctuates considerably over the period of use. Rayleigh fading model over log-normal shadowing is used. Rayleigh fading is most applicable when there is no dominant propagation along a line of sight between the transmitter and receiver. In Rayleigh fading environment, instantaneous power exhibits an exponential distribution (not in dB). Shadowing/fading (log-normal shadowing plus Rayleigh fading) effects are considered and Suzuki distribution is used. The mobile channels modeled by Suzuki process already include multipath propagation and shadowing [59].

RSS measurements in Watts obey an exponential distribution with mean received from log-normal distribution. Hence, the analysis in Watts is more appropriate in this case (shadowing/fading channel). The ideal RSS at the ith MCD, denoted by P_i^{rx}, is expressed as

$$P_i^{rx}[\text{Watts}] = E_i(10^{\frac{\phi}{10}}) \tag{3.48}$$

$$\phi = P^{tx}[\text{dB}] - P_{L_0} - 10\log_{10} d_{(x_i,y_i)}^{\alpha} + S_i$$

where $E_i(\mu)$ is an exponential random variable with mean μ and S_i is a Gaussian random variable with mean zero and variance σ_s^2 for expressing the effect of log-normal shadowing.

Since Rayleigh fading and shadowing cause statistically independent multiplicative fluctuations of the received power, attenuation in dB-values caused by shadowing and Rayleigh fading can be added. The results indicate that if logarithmic moments of the approximate log-normal distribution would be matched to the exact Suzuki distribution, the mean would be found to be 2.5 dB below the local-mean caused by shadowing [59]. Hence to reduce the fading effects, the sample mean of measurements are evaluated and the loss due to short term fading is considered. The mean power in this case is

$$\overline{P_i^{rx}}[\text{dB}] = \sum_{j=1}^{N_m} P_{i,j}^{rx}[\text{dB}]/N_m + 2.5[\text{dB}] \qquad (3.49)$$

where N_m is the total number of samples for unit measurement, $P_{i,j}^{rx}$ is the jth measured RSS value at the ith MCD in Watts and follows an exponential distribution. The RSS value from Exponential distribution in Watts is derived and converted them in dB to evaluate the mean in dB scale.

Channel model affects the RSS measurements and it depends on the coefficients α and σ_s. These parameters are modeled for some environments and situations via extensive measurements and data analysis. Therefore, utilizing these estimated parameters enhances the quality of constructed REM significantly [60].

3.8.4 REM Construction Techniques

REM construction is a wide concept referring to creating a complete map of CRN coverage area. However, we focus only on deriving interference level at each point of the CRN coverage area. This is also referred to as Interference Cartography (IC). IC forms a map of signal strength of RF environment as a function of spatial coordinates in a predefined area, measurements from CRs, and possibly time (for a dynamic environment).

Typically, a REM construction method performs the following tasks: data gathering, data processing to interpret the underlying model and deciding on the state of each pixel. If data is collected only internally from the nodes in CRN, REM may not be sufficiently accurate. Additionally, as REM construction costs time and power on CR nodes, it may become a burden on the network. Therefore, REM construction can be delegated to external entities such as sensor nodes dedicated to this task. In this case, CR nodes can also contribute to the environment modeling in their locality and fine tune the environment model based on their experience.

Some construction mechanisms may utilize propagation modeling, which is determined by the terrain properties between the transmitter and the location. If not

modeled appropriately, a significant portion of the spectrum may be wasted due to false alarms while it may also lead to miss detections. Therefore, REM construction methods should retrieve terrain data stored internally or externally, and compute the spectrum availability using these terrain data.

In the literature, REM construction techniques can be broadly put into two classes: spatial statistics based methods [44] and transmitter location determination based methods [3], which are also referred to as direct and indirect methods, respectively [13].

3.8.5   Spatial Statistics Based Methods

Spatial statistics describes the statistical properties of a given area utilizing the spatial correlational structure of this region. Using spatial statistics and given the measurements at specific locations, missing data at areas without any measurements can be estimated as a function of measured data. The fusion of the data from the examined area is commonly based on different types of interpolation methods. Interpolation is based on the basic principle that geographically close locations are more related to each other compared to the more distant locations. Possible interpolation techniques in the literature are Kriging, Inverse Distance Weighted (IDW), and nearest neighbor interpolation. The frequency or volume of measurements, required accuracy in measurements, and density of measurement point are key factors determining the performance of each method. Although Kriging requires more measurement points, it is the most commonly applied technique in the literature [24, 48] due to its higher precision.

Specific fields of spatial statistics are random fields and point processes. Random fields model the phenomena under study as a continuous region while point processes characterizes the locations of *events*. In a network context, point processes can be used to describe the location distribution of transmitters and receivers. Stochastic geometry of the wireless nodes is crucial to be represented in the design of network protocols. As PU density determines the interference, point processes are useful for the REM construction.

Instead of storing the raw data of locations, point processes describing the orientation of the nodes in the network with a number of parameters can be stored in the REM [44]. This descriptive and compact model reduces the storage requirements which is desirable especially for wide area REM operation. Additionally, compact representation of the system reduces the overhead of information exchange among different REMs.

3.8.5.1   Inverse Distance Weighted Interpolation

Inverse distance weighted interpolation basically gives weights to the measurements according to the inverse of the distance to the point of interest. A general form of

finding an interpolated value z at a given point p based on measurements $z_i = z(p_i)$ for $i = 1, 2, 3, \ldots, N_s$ using IDW-β is an interpolating function:

$$z(p) = \sum_{i=1}^{N_s} \frac{w_i(p)z_i}{\sum_{j=1}^{N_s} w_j(p)} \tag{3.50}$$

where N_s is the number of measurement points,

$$w_i(p) = \frac{1}{d(p, p_i)^\beta} \tag{3.51}$$

is a simple IDW weighting function of power β, and $d(., .)$ is a distance function.

Here, the weight decreases as the distance from the interpolated points increases. Larger values of β assign more influence to values closest to the interpolated point. For $0 < \beta < 1$, $z(p)$ has smooth peaks over the interpolated points p_i, while the peaks become sharp as $\beta > 1$. The choice of the value of β is, therefore, a function of the degree of smoothing desired in the interpolation, the density and distribution of samples being interpolated, and the maximum distance over which an individual sample is allowed to influence the surrounding points.

3.8.5.2 Kriging

Kriging is a group of geostatistical techniques that interpolate the value of a random field at an unobserved location from observations of its value at nearby locations. The theory behind interpolation and extrapolation by kriging was developed by the French mathematician Georges Matheron based on the Master's thesis of Daniel Gerhardus Krige [30], the pioneering plotter of distance-weighted average gold grades at the Witwatersrand reef complex in South Africa.

Kriging belongs to the family of linear least squares estimation algorithms. The aim of kriging is to estimate the value of an unknown real-valued function, f, at a point, x^*, given the values of the function at some other points, $x_1, \ldots, x_n$. A kriging estimator is said to be linear because the predicted value $\hat{f}(x^*)$ is a linear combination that may be written as

$$\hat{f}(x^*) = \sum_{i=1}^{n} \lambda_i(x^*) f(x_i). \tag{3.52}$$

The weights $\lambda_i(x^*)$ are solutions of a system of linear equations, which are obtained by assuming that f is a sample-path of a random process $F(x)$, and that the error of prediction

$$\varepsilon(x) = F(x) - \sum_{i=1}^{n} \lambda_i(x) F(x_i) \tag{3.53}$$

is to be minimized in some sense. For instance, the so-called simple kriging assumption is that the mean and the covariance of $F(x)$ is known and then, the kriging predictor is the one that minimizes the variance of the prediction error.

This method, originally used in mining exploration, has been applied in multitude of domains consisting of environmental science, meteorology, agriculture, and remote sensing. Central to geostatistics is the variogram, a function that models the variance between two points in space as a function of the distance between them. In the case of grid-sampled fields, the distance between measurements is a fixed lag distance. Randomized and optimized sampling schemes produce variable lag distances [41]. The weights used in linear combination depend on spatial correlation derived from *semivariogram* (half of the variogram) model of the measurement data. Semivariogram determines the correlation between any two points in the considered system based on their distance separation. In spatial statistics, the theoretical variogram $2\gamma(x, y)$ is a function describing the degree of spatial dependence of a spatial random field or stochastic process $Z(x)$. It is defined as the variance of the difference between field values at two locations across realizations of the field [12].

$$2\gamma(x, y) = \text{var}(Z(x) - Z(y)) = E\left(|(Z(x) - \mu(x)) - (Z(y) - \mu(y))|^2\right). \tag{3.54}$$

If the spatial random field has constant mean μ, this is equivalent to the expectation for the squared increment of the values between locations x and y [57]:

$$2\gamma(x, y) = E\left(|Z(x) - Z(y)|^2\right), \tag{3.55}$$

where $\gamma(x, y)$ itself is called the semivariogram.

Kriging interpolates the value $Z(x_0)$ of a random field $Z(x)$ at an unobserved location x_0 from observations $z_i = Z(x_i)$, $i = 1, \ldots, n$ of the random field at nearby locations $x_1, \ldots, x_n$. Kriging computes the best linear unbiased estimator $\hat{Z}(x_0)$ of $Z(x_0)$ based on a stochastic model of the spatial dependence quantified either by the variogram $\gamma(x, y)$ or by expectation $\mu(x) = E[Z(x)]$ and the covariance function $c(x, y)$ of the random field [10, 52].

Therefore, more formally, Kriging estimator is given by a linear combination

$$\hat{Z}(x_0) = \sum_{i=1}^{n} w_i(x_0) Z(x_i) \tag{3.56}$$

of the observed values $z_i = Z(x_i)$ with weights $w_i(x_0)$, $i = 1, \ldots, n$ chosen such that the variance (also called Kriging variance or Kriging error):

$$\sigma_k^2(x_0) \triangleq \text{Var}\left(\hat{Z}(x_0) - Z(x_0)\right) \tag{3.57}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i(x_0) w_j(x_0) c(x_i, x_j) + \text{Var}\left(Z(x_0)\right) - 2\sum_{i=1}^{n} w_i(x_0) c(x_i, x_0)$$

is minimized subject to the unbiasedness condition:

$$\mathrm{E}[\hat{Z}(x) - Z(x)] = \sum_{i=1}^{n} w_i(x_0)\mu(x_i) - \mu(x_0) = 0 \qquad (3.58)$$

The Kriging variance must not be confused with the variance

$$\mathrm{Var}\left(\hat{Z}(x_0)\right) = \mathrm{Var}\left(\sum_{i=1}^{n} w_i Z(x_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j c(x_i, x_j) \qquad (3.59)$$

of the Kriging predictor $\hat{Z}(x_0)$ itself.

The interpolation by simple Kriging is given by:

$$\hat{Z}(x_0) = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}^T \begin{pmatrix} c(x_1, x_1) & \cdots & c(x_1, x_n) \\ \vdots & \ddots & \vdots \\ c(x_n, x_1) & \cdots & c(x_n, x_n) \end{pmatrix}^{-1} \begin{pmatrix} c(x_1, x_0) \\ \vdots \\ c(x_n, x_0) \end{pmatrix} \qquad (3.60)$$

3.8.6 Transmitter Location Determination Based Methods

Aforementioned statistical methods directly approximate the signal strength without being concerned about the sources of the power. However, if information on PU locations are already available or can be approximated, they ease the process of REM construction. Transmitter location determination based methods use this approach and first focus on localizing the transmitter(s) and deriving their properties. Subsequently, they estimate the signal strength at each location by applying the propagation modeling. However, this approach has more degrees of freedom: multiple transmitters, transmitter properties such as antenna propagation pattern, and accurate characterization of the propagation environment.

After the potential transmitters are located, appropriate modeling of the signal strength calls for appropriate propagation modeling, e.g., channel gains between each pair of transmitter and receiver. In [44], Riihijärvi et al. study the effect of transmitter properties on the signal strength in a CRN using the second order statistics. In [3], authors applies an image processing based technique, which identifies the transmitters in the system and estimates their parameters based on Received Signal Strength (RSS) from sensors.

If you can estimate some channel parameters then using them for transmitter location estimation and the propagation of the signal improves the REM quality significantly [60].

3.9   Conclusion

This chapter covers cooperative spectrum sensing in a holistic manner. First, we discuss how the observations from numerous sensing devices can be utilized to maximize the probability of detection while keeping the probability of false alarms bounded. Due to different channel conditions, the observations of different SUs are weighted accordingly in the decision process.

Given a cooperative sensing algorithm, a wide spectrum to sense, and a set of sensing devices, one of the critical problems is to decide which device senses which spectrum band. Since the number of spectrum bands to be sensed is large and a number of devices must have performed sensing to enable cooperation, it is vital to correctly decide which bands should be sensed by which devices according to the current channel conditions. The second part of the chapter discusses spectrum sensing scheduling to address this issue.

Finally, we discuss REM as an enabler for the cognitive engine. REM constructs a multi-layer temperature map for all spectrum bands of interest. For each spectrum band, the region where SUs should abstain from utilizing that specific band is determined. REM is constructed by utilizing cooperative spectrum sensing techniques on the observations of the SUs which were selected by the sensing scheduling method. The outcome of REM can be used to decide which spectrum band can be utilized by the SUs at a specific location.

References

1. Arslan, H.: Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems. Springer, Dordrecht (2007)
2. Atapattu, S., Tellambura, C., Jiang, H.: Energy detection based cooperative spectrum sensing in cognitive radio networks. IEEE Trans. Wirel. Commun. **10**(4), 1232–1241 (2011)
3. Bolea, L., Pérez-Romero, J., Agustí, R., Sallent, O.: Context discovery mechanisms for cognitive radio. In: Proceedings of the 73rd Vehicular Technology Conference (VTC), Budapest, pp. 1–5 (2011)
4. Cabric, D., Mishra, S., Brodersen, R.: Implementation issues in spectrum sensing for cognitive radios. In: Conference Record of the 38th Asilomar Conference on Signals, Systems and Computers, vol. 1, pp. 772–776, Pacific Grove (2004)
5. Cabric, D., Tkachenko, A., Brodersen, R.: Spectrum sensing measurements of pilot, energy, and collaborative detection. In: Proceedings of the Military Communications Conference (MILCOM), pp. 1–7, Washington, D.C. (2006)
6. Cattoni, A.F., Minetti, I., Gandetto, M., Niu, R., Varshney, P.K., Regazzoni, C.S.: A spectrum sensing algorithm based on distributed cognitive models. In: Proceedings of the SDR Forum Technical Conference, Orlando (2006)
7. Chair, Z., Varshney, P.K.: Optimal data fusion in multiple sensor detection systems. IEEE Transact. Aerosp. Electron. Syst. **AES-22**(1), 98–101 (1986)

8. Chaudhari, S., Lundén, J., Koivunen, V.: Bep walls for collaborative spectrum sensing. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, pp. 2984–2987 (2011)

9. Chen, H., Chen, H.H.: Spectrum sensing scheduling for group spectrum sharing in cognitive radio networks. Wiley Int. J. Commun. Syst. **24**(1), 62–74 (2011)

10. Chiles, J.P., Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty. Wiley, New York (2012)

11. Conejo, A., Castillo, E., Minguez, R., Garcia-Bertrand, R.: Decomposition Techniques in Mathematical Programming. Springer, Berlin/New York (2006)

12. Cressie, N.: Statistics for spatial data. Wiley Terra Nova **4**(5), 613–617 (1992)

13. Dagres, I., Polydoros, A., Riihijärvi, J., Nasreddine, J., Mähönen, P., Gavrilovska, L., Atanasovski, V., van de Beek, J., Sayrac, B., Grimoud, S., Benitez, M.L., Romero, J.P., Agusti, R., Casadevall, F.: Flexible and spectrum-aware radio access through measurements and modelling in cognitive radio systems faramir, D4.1 radio environmental maps: information models and reference model. Technical report FARAMIR EU Project (2011)

14. Digham, F.F., Alouini, M.S., Simon, M.K.: On the energy detection of unknown signals over fading channels. In: Proceedings of the International Conference on Communications (ICC), Anchorage, vol. 5, pp. 3575–3579 (2003)

15. Digham, F.F., Alouini, M.S., Simon, M.K.: On the energy detection of unknown signals over fading channels. IEEE Trans. Commun. **55**(1), 21–24 (2007)

16. Eryigit, S., Bayhan, S., Tugcu, T.: Channel switching cost aware and energy-efficient cooperative sensing scheduling for cognitive radio networks. In: IEEE International Conference on Communications (ICC), Dresden, pp. 2633–2638. IEEE (2013): ©[2013] IEEE. Reprinted, with permission, from Eryigit, S., Bayhan, S., Tugcu, T.: Channel switching cost aware and energy-efficient cooperative sensing scheduling for cognitive radio networks. In: Proceedings of IEEE International Conference on Communications, Dresden, June 2013

17. Eryigit, S., Bayhan, S., Tugcu, T.: Energy-efficient multichannel cooperative sensing scheduling with heterogeneous channel conditions for cognitive radio networks. IEEE Trans. Veh. Technol. **62**(6), 2690–2699 (2013). doi:10.1109/TVT.2013.2247070

18. Gandetto, M., Cattoni, A.F., Regazzoni, C.S.: A distributed approach to mode identification and spectrum monitoring for cognitive radios. In: Proceedings of the SDR Forum Technical Conference, Anaheim (2005)

19. Gandetto, M., Regazzoni, C.S.: Spectrum sensing: a distributed approach for cognitive terminals. IEEE J. Sel. Areas Commun. **25**(3), 546–557 (2007)

20. Ganesan, G., Li, Y.: Agility improvement through cooperative diversity in cognitive radio. In: Proceedings of the Global Telecommunications Conference (GLOBECOM), Missouri, vol. 5, pp. 2505–2509 (2005)

21. Ganesan, G., Li, Y.: Cooperative spectrum sensing in cognitive radio networks. In: Proceedings of the First International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, pp. 137–143 (2005)

22. Geirhofer, S., Tong, L., Sadler, B.: A measurement-based model for dynamic spectrum access in wlan channels. In: Proceedings of the Military Communications Conference (MILCOM), Washington, D.C., pp. 1–7 (2006)

23. Gozupek, D., Buhari, S., Alagoz, F.: A spectrum switching delay aware scheduling algorithm for centralized cognitive radio networks. IEEE Trans. Mob. Comput. **12**(7), 1270–1280 (2013)

24. Grimoud, S., Ben Jemaa, S., Sayrac, B., Moulines, E.: A REM enabled soft frequency reuse scheme. In: Proceedings of the GLOBECOM Workshops (GC Wkshps), Miami, pp. 819–823 (2010)

25. Hao, X., Cheung, M., Wong, V., Leung, V.: A coalition formation game for energy-efficient cooperative spectrum sensing in cognitive radio networks with multiple channels. In: IEEE Global Telecommunications Conference (GLOBECOM 2011), Houston, pp. 1–6 (2011)

26. Harrison, K., Mishra, S.M., Sahai, A.: How much white-space capacity is there? In: Proceedings of the International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Singapore, pp. 1–10 (2010)

27. Hashemi, H.: The indoor radio propagation channel. Proc. IEEE **81**(7), 943–968 (1993)

28. Herath, S., Rajatheva, N., Tellambura, C.: Energy detection of unknown signals in fading and diversity reception. IEEE Trans. Commun. **59**(9), 2443–2453 (2011)
29. Kaligineedi, P., Bhargava, V.K.: Sensor allocation and quantization schemes for multi-band cognitive radio cooperative sensing system. IEEE Trans. Wirel. Commun. **10**(1), 284–293 (2011)
30. Krige, D.G.: A statistical approach to some mine valuations and allied problems at the witwatersrand. Master's thesis, University of Witwatersrand (1951)
31. Lee, S.H., Oh, D.C., Lee, Y.H.: Hard decision combining-based cooperative spectrum sensing in cognitive radio systems. In: Proceedings of the International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly, Leipzig, pp. 906–910 (2009)
32. Leu, A.E., McHenry, M., Mark, B.L.: Modeling and analysis of interference in listen-before-talk spectrum access schemes. Wiley Int. J. Netw. Manag. **16**(2), 131–147 (2006)
33. Liang, Y., Zeng, Y., Peh, E., Hoang, A.: Sensing-throughput tradeoff for cognitive radio networks. IEEE Trans. Wirel. Commun. **7**(4), 1326–1337 (2008)
34. Mishra, S.M., Sahai, A., Brodersen, R.W.: Cooperative sensing among cognitive radios. In: Proceedings of the International Conference on Communications (ICC), Istanbul, vol. 4, pp. 1658–1663 (2006)
35. Morgado, A., Carvalho, N.B.: White spaces communications in Europe. In: Proceedings of the 30th General Assembly and Scientific Symposium (URSI), Istanbul, pp. 1–4 (2011)
36. Murty, R., Chandra, R., Moscibroda, T., Bahl, P.V.: Senseless: a database-driven white spaces network. IEEE Trans. Mob. Comput. **11**(2), 189–203 (2012)
37. Nguyen-Thanh, N., Koo, I.: Evidence theory based cooperative spectrum sensing with efficient quantization method in cognitive radio. IEEE Trans. Veh. Technol. **60**(1), 185–195 (2011)
38. Oh, D.C., Lee, Y.H.: Cooperative spectrum sensing with imperfect feedback channel in the cognitive radio systems. Wiley Int. J. Commun. Syst. **23**(6–7), 763–779 (2010)
39. Pawełczak, P., Guo, C., Prasad, R.V., Hekmat, R.: IRCTR-S-004-07: cluster-based spectrum sensing architecture for opportunistic spectrum access networks. Technical report, International Research Centre for Telecommunications and Radar (2006)
40. Peh, E., Liang, Y.C.: Optimization for cooperative sensing in cognitive radio networks. In: Proceedings of the Wireless Communications and Networking Conference (WCNC), Hong Kong, pp. 27–32 (2007)
41. Phillips, C., Ton, M., Sicker, D., Grunwald, D.: Practical radio environment mapping with geostatistics. In: Proceedings of the Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Bellevue (2012)
42. Picinbono, B., Duvaut, P.: Optimum quantization for detection. IEEE Trans. Commun. **36**(11), 1254–1258 (1988)
43. Qihang, P., Kun, Z., Jun, W., Shaoqian, L.: A distributed spectrum sensing scheme based on credibility and evidence theory in cognitive radio context. In: Proceedings of the 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsiniki, pp. 1–5 (2006)
44. Riihijärvi, J., Mähönen, P., Sajjad, S.: Influence of transmitter configurations on spatial statistics of radio environment maps. In: Proceedings of the 20th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Tokyo, pp. 853–857 (2009)
45. Sakran, H., Shokair, M.: Hard and softened combination for cooperative spectrum sensing over imperfect channels in cognitive radio networks. Springer Telecommun. Syst. **52**(1), 61–71 (2013)
46. Sakran, H., Shokair, M., El-Rabaie, E.S., El-Azm, A.A.: Three bits softened decision scheme in cooperative spectrum sensing among cognitive radio networks. In: Proceedings of the 28th National Radio Science Conference (NRSC), Delhi, pp. 1–9 (2011)
47. Shankar, N., Cordeiro, C., Challapali, K.: Spectrum agile radios: utilization and sensing architectures. In: Proceedings of the First International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, pp. 160–169 (2005)

48. Subramani, S., Riihijarvi, J., Sayrac, B., Gavrilovska, L., Sooriyabandara, M., Farnham, T., Mahonen, P.: Towards practical REM-based radio resource management. In: Proceedings of the Future Network and Mobile Summit (FutureNetw), Warsaw, pp. 1–8 (2011)
49. Sun, C., Zhang, W., Letaief, K.B.: Cooperative spectrum sensing for cognitive radios under bandwidth constraints. In: Proceedings of the Wireless Communications and Networking Conference (WCNC), Hong Kong, pp. 1–5 (2007)
50. Sun, X., Zhang, T., Tsang, D.: Optimal energy-efficient cooperative sensing scheduling for cognitive radio networks with qos guarantee. In: 7th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), Istanbul, pp. 1825–1830 (2011)
51. Tandra, R., Sahai, A.: Fundamental limits on detection in low SNR under noise uncertainty. In: Proceedings of the International Conference on Wireless Networks, Communications and Mobile Computing (WirelessCom), Maui, vol. 1, pp. 464–469 (2005)
52. Tonkin, M.J., Larson, S.P.: Kriging water levels with a regional-linear and point-logarithmic drift. Wiley Gr. Water **40**(2), 185–193 (2002)
53. UBAK: Türkiye'de dvb-t. http://www.ubak.gov.tr/BLSM_WIYS/HGB/tr/Sag_Menu/20100816_162441_10472_1_64.html. Accessed at May 2011
54. Unnikrishnan, J., Veeravalli, V.: Cooperative sensing for primary detection in cognitive radio. IEEE J. Sel. Top. Signal Process. **2**(1), 18–27 (2008)
55. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. IEEE **55**(4), 523–531 (1967)
56. Visotsky, E., Kuffner, S., Peterson, R.: On collaborative detection of TV transmissions in support of dynamic spectrum sharing. In: Proceedings of the First International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, pp. 338–345 (2005)
57. Wackernagel, H.: Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin (2003)
58. Weiss, T.A.: A diversity approach for the detection of idle spectral resources in spectrum pooling systems. In: Proceedings of the 48th International Scientific Colloquium, Ilmenau, pp. 37–38 (2003)
59. Wirastuti, N., Sastra, N.P.: Application of the suzuki distribution to simulations of shadowing/fading effects in mobile communication. In: Proceedings of the Fourth International Conference on Information and Communication Technology and System, Surabaya Indonesia (2008)
60. Yilmaz, H.B., Tugcu, T.: Location estimation-based radio environment map construction in fading channels. Wirel. Commun. Mob. Comput. (2013). doi:10.1002/wcm.2367
61. Yilmaz, H.B., Tugcu, T., Alagoz, F.: Uniform quantizer for cooperative sensing in cognitive radio networks. In: Proceedings of the 21st International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, pp. 548–553 (2010)
62. Yuan, Y., Bahl, P., Chandra, R., Chou, P.A., Ferrell, J.I., Moscibroda, T., Narlanka, S., Wu, Y.: Knows: Cognitive radio networks over white spaces. In: Proceedings of the Second International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Dublin, pp. 416–427 (2007)
63. Zhang, T., Tsang, D.: Optimal cooperative sensing scheduling for energy-efficient cognitive radio networks. In: IEEE International Conference on Computer Communications (INFO-COM), pp. 2723–2731 (2011)
64. Zhang, W., Mallik, R., Letaief, K.: Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. IEEE Trans. Wirel. Commun. **8**(12), 5761–5766 (2009)
65. Zhang, X., Wu, Q., Wang, J.: Optimization of sensing time in multichannel sequential sensing for cognitive radio. Wiley Int. J. Commun. Syst. (2011). doi:10.1002/dac.1341
66. Zhao, Y., Le, B., Reed, J.H.: Network support–the radio environment map. In: Cognitive Radio Technology, pp. 337–363. Elsevier, Amsterdam/Boston (2006)

Chapter 4
Medium Access Control Protocols in Cognitive Radio Networks

**Liljana Gavrilovska, Daniel Denkovski, Valentin Rakovic,
and Marko Angjelicinoski**

Abstract The endeavor of categorizing the existing Cognitive-MAC (C-MAC) protocols requires definition of general classification frameworks or layouts that merge most of the aspects of the protocols in a single unified presentation. This chapter introduces the C-MAC cycle as a general classification and systematization layout for C-MAC protocols. The C-MAC cycle originates form the idea that the MAC layer in spectrally heterogeneous environments should provide support for three generic technical features: radio environmental data acquisition; spectrum sharing; and control channel management. The inclusion of these generic technical features is necessary in Cognitive Radio Networks (CRNs) for improving the network performance and achieving spectrum efficiency gain while providing maximal level of protection for the primary system. This chapter presents extensive survey on the state-of-art advances in C-MAC protocol engineering by reviewing existing technical solutions and proposals, identifying their basic characteristics and placing them into the C-MAC cycle, with emphasis on the modularity of the C-MAC cycle. It provides overview of large number of technical details concerning the three generic functionalities (i.e. the radio environmental data acquisition, the spectrum sharing and the control channel management) as the main building blocks of the C-MAC cycle. Three uses cases (each in different generic functional group), illustrate the capabilities of the proposed C-MAC cycle layout. In more detail, the first use case theoretically presents and practically evaluates cooperative spectrum sensing based on Estimated Noise Power. The results illustrate the effect of estimating the noise variance on the detection capabilities of the Majority Voting and Equal Gain Combining cooperative spectrum sensing strategies. The second

L. Gavrilovska (✉) • D. Denkovski • V. Rakovic • M. Angjelicinoski
Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius
University, Skopje, Macedonia
e-mail: liljana@feit.ukim.edu.mk; danield@feit.ukim.edu.mk; valentin@feit.ukim.edu.mk;
markoang@feit.ukim.edu.mk

use case presents advanced and computationally efficient horizontal spectrum sharing strategy for secondary systems based on Node Clustering and Beamforming. Finally, the last use case presents and assesses a multiuser quorum-based multiple rendezvous strategy for control channel establishment in distributed Cognitive Radio Networks.

4.1 Introduction

The scientific community has recently experienced rapidly increasing interest in definition and design of Medium Access Control protocols for Cognitive Radio Networks (CRNs). As witnessed, the concept of Cognitive Radio (CR) and CRNs evolved from an abstract and general idea and broad theoretical topic [1] to a set of applicable and practically deployable solutions, some of which entered the process of standardization [2, 3]. These standardization efforts mostly rely on the Dynamic Spectrum Access (DSA) concept (which achieved the status of synonym for CRN) and most networking solutions related to CRN standardization focus on enabling efficient operation in opportunistic i.e. DSA environments. However, the CR as introduced in [1] is envisioned as versatile and dynamically adaptable technology enabler, capable of learning and reasoning possessing the key potential for improving the efficiency and quality of the wireless communication in various different environments and operational settings.

The main distinctive characteristic considering the radio environmental conditions where CRNs (i.e. the secondary systems) operate is the variable availability of the spectrum resources in time, space and frequency [4], a phenomenon commonly referred as **spectrum heterogeneity**. Its behavior and characteristics are the main factors that determine the secondary system performance, the level of protection of the licensed wireless network (i.e. the primary system) and the overall spectrum efficiency gain. The spectrum heterogeneity imposes redefinition of the protocol stack (especially the lower layers i.e. the physical, Medium Access Control and network layer) by introducing new communication protocols. Additionally, cross layering and tight operation coupling between the layers of the stack is highly recommended to achieve better secondary system performances. The novel cross layering-enabled communication protocols address the spectrum heterogeneity by primary system behavior monitoring and transparent adapting to the variable spectrum availability and network topology.

This chapter focuses on MAC protocols designed for CRNs i.e. **Cognitive-MAC (C-MAC)** protocols. In particular, it emphasizes their overall importance for efficient operation of CRN, highlights their basic features with respect to the operational settings and identifies the main challenges concerning the design of C-MAC protocols that enable large spectrum efficiency gains. Summarizing the recent research achievements [4–7], each C-MAC protocol that efficiently addresses the spectrum heterogeneity, should provide support for the following functional requirements [8]:

1. *Primary system protection*. The secondary system should provide maximal protection for the primary system. It should not disrupt the communication links of the primary users by defining strategies for secondary-to-primary interference avoidance and mitigation for attaining the required transparency.
2. *Access to radio environmental information*. The radio environmental knowledge should be acquired by enabling high fidelity spectrum sensing mechanisms and/or access to up-to-date radio environmental information stored in databases. This radio context information should serve as the main enabler of radio environmental awareness.
3. *Advanced spectrum sharing techniques*. C-MAC is responsible for enabling efficient and dynamic spectrum access and resource allocation which aims to increase the overall performance of secondary system by exploiting advanced and intelligent spectrum sharing techniques. Additionally, efficient spectrum sharing strategies should also serve as a facilitating tool for primary system protection.
4. *Control signaling mechanisms*. Fully operational C-MAC protocol requires efficient management and reliable dissemination of control data through identification, definition, establishment and management of reliable and secure control channel.

Addressing and meeting these functional requirements is of high importance towards increasing the overall spectrum efficiency by smart exploitation of its current underutilized space/time/frequency regions. Plethora of proposals, strategies and fully operational and efficient C-MAC solution has been reported in the literature. Most of these proposals address some subset of the following implementation challenges and research topics related to the aforesaid functional requirements: cooperative spectrum sensing, multiband operation, coordination among network nodes, spectrum access and allocation by exploiting advanced artificial intelligence and optimization techniques, secondary-to-primary but also primary-to-secondary interference mitigation and avoidance, multi-channel hidden terminal problem resolving, control channel configuration and reconfiguration, mobility, security, QoS support, scheduling, ARQ procedures etc. Traditionally, the distinction between the different MAC protocols, their classification and systematization is performed on a basis of the employed medium access scheme [9]. However, such practice in the case of C-MAC protocols is practically impossible due to the multidimensional nature of the C-MAC protocols which try to address and meet several conflicting and contradictory requirements as a result of spectrum heterogeneity. The classification of existing C-MAC protocols requires definition of general classification frameworks or layouts that merge most of the aspects of the protocols in a single unified presentation. Such generic layout should provide firm understanding of the protocol operation and it should serve as facilitating tool for the future C-MAC protocol engineering process.

This chapter presents the C-MAC cycle as a general classification and systematization layout for C-MAC protocols [8]. The C-MAC cycle originates form the idea that the MAC layer in spectrally heterogeneous environments should provide

support for three generic functionalities: radio environmental data acquisition and knowledge building for radio environmental awareness; efficient radio resource sharing i.e. spectrum sharing; and efficient control channel management. The chapter also provides brief survey on the state-of-art advances in C-MAC protocol engineering by reviewing existing technical solutions and proposals, identifying their basic characteristics and placing them into the C-MAC cycle, with emphasis on the modularity of the C-MAC cycle. It gives an overview of large number of technical details concerning the three generic functionalities and the related functionality-specific and common aspects. Three uses cases (each in different generic functional group), stemmed from authors' previous experience in the area, illustrate the capabilities of the proposed C-MAC cycle layout. In more detail, the first use case theoretically presents and practically evaluates cooperative spectrum sensing based on Estimated Noise Power. The results illustrate the effect of estimating the noise variance on the detection capabilities of the Majority Voting and Equal Gain Combining cooperative spectrum sensing strategies. The second use case presents advanced and computationally efficient horizontal spectrum sharing strategy for secondary systems based on Node Clustering and Beamforming. Finally, the last use case presents and assesses a multiuser quorum-based multiple rendezvous strategy for control channel establishment in distributed Cognitive Radio Networks.

The rest of the chapter is organized as follows. Section 4.2 summarizes the existing C-MAC classification and systematization attempts, highlighting their major drawbacks. Section 4.3 presents the structure and the main motivation behind the C-MAC cycle as generic C-MAC protocol classification layout. Section 4.4 briefly overviews the three generic functionalities of the C-MAC cycle i.e. the spectrum sensing, sharing and control channel management. Section 4.5 presents the C-MAC cycle use cases. Section 4.6 concludes the chapter.

4.2 C-MAC Protocol Classification and Systematization

Several C-MAC protocol classification attempts have been reported in the literature [10–13]. They reflect the state-of-the-art research and standardization achievements in CR networking, differentiate MAC protocols designed for CRNs from MAC protocols designed for legacy wireless networks and attempt to identify general C-MAC classification and systematization criteria. Table 4.1 summarizes the reported criteria used for C-MAC classification in recent publications.

There are several major drawbacks and limitations in these classification proposals. In particular, they generally focus on and tend to cover only some specific set of C-MAC features. Focusing on a subset of C-MAC protocol features and rendering all the existing work on the topic through them, while partially or completely circumventing and ignoring other equally important aspects, results in loss of generality and creation of confusing semantics.

Table 4.1 Proposed C-MAC classification criteria [10–13]

According to	Solutions	Classification based on
Network infrastructure (architecture based)	Centralized or distributed	Location of spectrum management entity
Spectrum resource management	Centralized spectrum access networks or centralized spectrum allocation networks	Role of centralized network entity (only for centralized approaches)
Control channel establishment	Global, local, configurable, w/o dedicated control channel	Presence, scope and characteristics of the control channel (only for distributed approaches)
Spectrum sensing technique	Local or cooperative	Degree of mutual nodes' interactions
Spectrum access modes	Contention-based, time-slotted or hybrid	Secondary nodes spectrum access
Spectrum usage strategy	Single channel or multi-channel	Number of channels available to the secondary nodes
Spectrum sharing modes	Overlay, underlay or interweave	Degree of cooperation between primary and secondary
Number of radios	Single radio or multi radio	Number of radios available to the secondary nodes
Optimization and learning (network coordination)	Direct Access Based (DAB) protocols or Dynamic Spectrum Allocation (DSA) protocols	Degree of optimization scope

The existing literature does not provide systematization and classification layout with unified presentation of all generic (*fundamental*) and *optional* functionalities supported by efficient and reliable C-MAC protocols. The next section introduces the C-MAC cycle, designed to alleviate this deficiency.

4.3 Generic C-MAC Protocols Layout: The C-MAC Cycle

A general C-MAC classification layout should address the following requirements: it should provide firm understanding of the protocol operation; it should be modular, flexible and easily extensible; and it should serve as a future protocol design facilitating tool. The proposed C-MAC cycle [8] meets all of these requirements. The main underlying idea that generates the concept of C-MAC cycle is the recognition that a single C-MAC protocol should support and implement at least three *generic* functionalities for efficient operation in spectrally heterogeneous environment (Fig. 4.1). These generic functionalities are: *radio environment data acquisition*, the *spectrum sharing* and the *control channel management*. The requirement for their mandatory support distinguishes C-MAC protocols from the common legacy MAC protocols.
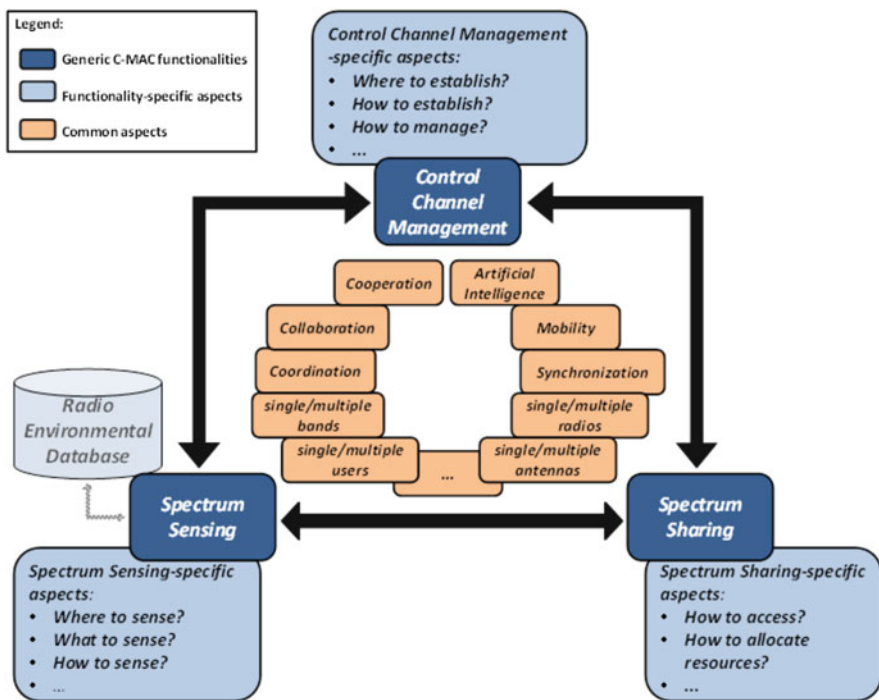
Fig. 4.1 C-MAC cycle [8]

Each of the three generic functionalities is associated to several ***functionality specific aspects*** (i.e. issues that each of the generic functionalities encounters and tends to solve) and common features, techniques and mechanisms referred as ***common aspects***. The distinction between common and functionality specific aspects can significantly improve the flexibility and modularity of the C-MAC cycle. The functionality specific aspects impose challenges and limitations that might be addressed and resolved by using some common techniques and strategies represented by the common aspects. Other than that, the set of common aspects also includes some major limitations imposed on the C-MAC protocols that usually relate to the CR hardware configuration or the operational mode of the network.

The next section overviews the three generic functionalities of the C-MAC cycle, covers their most important functionality specific aspects and discusses the extent to which the common aspects can be utilized to address the functionality specific issues considering the operational settings of the network. The importance of various common aspects (such as optimization, cooperation, coordination among secondary users) for efficient design of C-MAC protocols is specifically highlighted and elaborated throughout the rest of the chapter.

4.4 Overview of the Generic C-MAC Functionalities

This section aims to briefly describe the three generic functionalities of the C-MAC functionalities as well the common techniques and strategies employed to address their specific aspect with respect to the operational settings of the secondary system.

4.4.1 Spectrum Sensing Strategies

There are two possible ways to obtain the radio environmental information and knowledge: via spectrum sensing (sensing-centric) or through an access to a radio environmental database (database-centric). The spectrum sensing is a physical layer functionality that is tightly coupled with the MAC layer and is perceived as the basic tool in CR for acquiring radio environmental data in the sensing-centric CR solutions. The data provided by the spectrum sensing functionality should provide reliable and up-to-date information on the spectrum occupancy, availability and usage. In the database-centric solutions, the radio context information is obtained through access to remote and (logically) centralized up-to-date environmental database. In this case the spectrum sensing functionality may be completely absent and the existence of a control channel mechanism for accessing and obtaining spectrum data from the central database, fulfils the radio environmental awareness requirement. Thus, when operating in database-centric mode, the spectrum sensing functionality is not mandatory. However, its implementation can be beneficial for improving the secondary system performance in general. Furthermore, enabling the spectrum sensing functionality is a challenging task that attracts attention by both, the industry and the research community. Therefore, this chapter focuses on the spectrum sensing functionality of the CMAC cycle.

4.4.1.1 Types of Spectrum Sensing

The main goal of the spectrum sensing functionality as a part of the C-MAC protocol is the discovery and the constant tracking of the spectrum opportunities for the operation of the CRN. The detection of spectrum opportunities can be performed via signal detection and classification. The ***signal detection techniques*** reported in the literature are classified in two broad classes: blind and feature detection techniques [14]. The *blind detection techniques* are used to blindly detect the presence or the absence of any type of signal in the wireless medium without any prior knowledge on the type and structure of the underlying primary user signals. Typical representatives of the class of blind detection techniques are the energy detection, the detection based on autocorrelation and the Higher-Order-Statistics detection [15]. The *feature detection* methods, in addition to the detection of signal existence, have the ability to perform signal classification, and hence, distinguish

primary from secondary users' signals, as well as distinguish between different types of primary and secondary user signals. Typical examples of feature detection techniques are the matched filter and cyclostationarity detection techniques. In *cooperative environments* [14, 16], individual spectrum observations can be fused into joint primary user signal existence decision by the means of using hard and soft ***decision fusion techniques***. In the case of *soft decision fusion*, the individual observation samples are summed using Equal Gain Combining (EGC), Maximum Ratio Combining (MRC) techniques, as well as other more advanced weighted based combining techniques [17], before making the decision on the signal presence. In *hard decision fusion* methods, such as AND, OR, M-out-of-N rules, Chair-Varshney fusion rule [18], individual decisions of the cooperating nodes are cooperatively fused into joint primary user decisions.

There are several general types of spectrum sensing depending on the spectrum sensing-specific aspect taken into account. With respect to the ***sensing execution time***, there are two types of spectrum sensing: reactive (on-demand) and proactive (periodic) sensing [2, 19]. The *reactive sensing* serves for spectrum opportunity discovery and it might be triggered by the radio environmental changes. The *proactive sensing* is used for persistent search and tracking of spectrum opportunities as well as estimating primary user activity patterns and traffic characteristics. Regarding the ***bands of interest*** for the secondary system, two types of sensing can be distinguished: in-band sensing and out-of-band sensing [2, 19]. The *in-band sensing* tracks for primary user signals and attempts to avoid harmful collisions on the same channel where secondary data transfer occurs. Oppositely, the *out-of-band sensing* aims to discover new opportunities for secondary usage. Based on the ***number of sensed bands***, the spectrum sensing is usually divided into two broad classes: single-band and multiband sensing [20]. In *single-band* mode, the spectrum sensing functionality senses and tends to discover secondary transmission opportunities in a single primary user band. On the other hand, when *multi-band sensing* is enabled, the spectrum sensing explores multiple legacy bands which results in increased flexibility and improved efficiency. However, the implementation of multi-band sensing is more complex and challenging.

4.4.1.2   Optimization of the Spectrum Sensing

In order to provide improved secondary system performances and additional spectrum efficiency gain, the spectrum sensing functionality should be optimized. In most cases, the spectrum sensing optimization goal is multi-objective resulting in multidimensional optimization problem with a number of (very often conflicting) constraints. The proposed strategies for sensing optimization usually rely on the common aspects of the C-MAC cycle and they can be classified (as shown on Table 4.2) in three general categories: strategies for single-band or multiple bands (non cooperative environments) and strategies for cooperative environments.

For **non-cooperative single-band** enabled environments, the common optimization parameters are the ***sensing period duration*** and the ***transmission period***

Table 4.2 Spectrum sensing optimization

Sensing optimization			
Non-cooperative environments			
Single-band	Multiple bands		Cooperative environments
Sensing time duration	Sensing Policies	Channel set Channel number Channel sensing order	Node selection
Transmission time duration	Stopping rule		Node clustering
Sensing errors	Exploration-exploitation trade-off		Channel selection

duration. The duration of the sensing period is tightly related to the reliability of the sensing, e.g. the probability of detection of a primary user signal and the probability of false alarm. However, the problem of sensing inefficiency arises with increasing sensing period duration. There exist a trade-off between the sensing and transmission periods' durations. Increasing the sensing period duration results in better sensing reliability but also in inefficient use of the transmission opportunities [21], while the increase of the transmission period duration increases the spectrum opportunity usage, but false alarms and miss-detections of primary user signals are more likely to occur. In addition to the sensing reliability related parameters, the optimization of the transmission and sensing periods durations, should also consider the primary system traffic behaviour. This can be done by adopting some predefined primary user channel model with fitted parameters such as the Gilber-Eliot primary user channel model [22].

In **multi-band** enabled environments, the spectrum sensing is optimized by deriving the optimal rules and means to resolve the set, the number and the order of the primary user channels to be sensed. These sets of rules are referred as spectrum sensing policies and they represent the main optimization target in multi-band enabled sensing.

With respect to the ***number of primary user bands*** covered by the spectrum sensing functionality, the sensing policies can be divided into two broad classes [20]. The first class refers to *single-channel sensing policies*, where the secondary user operating in a multi-band environment senses only a single legacy channel and based on the outcome (free or busy), the user decides whether to exploit the channel or to wait for other opportunities. The second class of the sensing policies, which is more challenging, relates to the sequential channel sensing. The CRN nodes adopting the *sequential-channel sensing policies* sense multiple primary user channels, before making the channel selection and access/sharing decisions. While the sequential-channel sensing provides dominant CRN performances, it also requires higher processing power.

In addition to selecting the sensing channel(s), in the case of sequential sensing policies, the ***number of legacy channels*** covered in a single C-MAC round is also crucial parameter [23, 24]. The performance of the sequential sensing depends on

the maximum number of channels a CR node can sense in a single C-MAC round. If the CR node is able to employ the *full observation sensing policy* i.e. the CR node is capable of performing the sensing of the full legacy band of interest, the best possible secondary system performance can be achieved. However, employing this sensing policy is not energy efficient. This issue pinpoint the need for more optimal and more efficient spectrum sensing policy (i.e. *green spectrum sensing policy*) [23, 25] that trades-off between the secondary system performance and the energy efficiency of the sensing process.

The **channel sensing order** [20, 26] is another important aspect covered by the sensing policies. The optimal channel sensing order selection among the contending secondary users yields a collision free secondary spectrum access. Various strategies and techniques for design of optimal channel sensing orders can be found in the existing literature. For an example, channel sensing orders can be selected randomly from all possible channel permutations, or they can be selected from a Latin square of non-overlapping channel permutations [20]. The sensing order selection process can be aided by learning and other advanced artificial intelligence concepts, efficiently reaching a collision free sensing orders based on the feedback from the spectrum sharing functionality.

Learning is a common C-MAC aspect that can be incorporated in the spectrum sensing policies to improve the performances of the generic C-MAC functionalities and the overall secondary system performances. With respect to the learning capabilities the spectrum sensing policies are divided into non-learning and learning policies. By using *non-learning policies* the CR node aims to perform the selection of sensing channels without considering historical data, such as previous availabilities and opportunities in the pool of legacy channels, as well as the previous outcomes (feedbacks) from the spectrum sharing functionality. Representatives of such non-learning policies are the random sensing policies [27, 28] and negotiation based sensing policies [27]. The *learning based sensing policies* [22, 29, 30] aim to improve the spectrum sensing process in the next C-MAC rounds by using historical data and experience accumulated in previous C-MAC rounds. The learning based sensing policies usually adopt a predefined primary system model and tend to learn the traffic parameters of the model and to do so, they exploit the historical sensing data, as well as the previous spectrum sharing feedbacks. The common primary user channel model extensively used in CR networking is the Gilber-Eliot channel model. Typical representative examples of the learning based policies are the myopic sensing [29, 30] and its variations.

The **sensing stopping rule** i.e. the rule that decides when to stop to sense the legacy channels, determines the overall duration of the sensing in multi-band environments and it's therefore a common optimization parameter. The *optimal sensing stopping rule* comes from the economics [31]. The secondary user compares the current reward with the expected reward if the sensing is continued, and stops as soon as the current reward is higher than the expected reward. The reward is a function of the channel observations and usually refers to some secondary system performance metric such as the aggregate secondary throughput calculated on the detected available channels. The current and the expected rewards are calculated

using some predefined primary user channel model. When the total number of sensing channels is K and the number of already sensed channels is n, then the current reward is calculated over these n sensed channels and the expected reward is calculated over all K channels including the remaining $K - n$ not sensed channels. This results in complex computations. A complexity reduction solution to the optimal stopping rule is the k-stage look-ahead rule. Instead of calculating the expected reward for all remaining sensing channels from the pool of K possible channels, the k-stage look-ahead rule compares the current reward with the expected reward over the subsequent $k < K$ channels. The authors in [31] show that there is only a slight secondary system performance decrease in the case of 2-stage look-ahead rule, compared to the optimal stopping rule case.

In learning aided multi-band spectrum sensing scenarios, a common trade-off that arises is the *exploration-exploitation trade-off* [22, 26]. The secondary user might decide to probe and access the already proven most reliable channels, or it can try to explore new arisen opportunities. This trade-off highly affects the secondary system performances in the cognitive radio environment. A common methodology in the learning aided spectrum sensing and access is to perform the sensing channels selection based on the exploration-exploitation trade-off. Thus, the channels to be sensed can be obtained as an outcome of optimization. The optimization function is usually composed of an exploitation part, which prefers historically most available channels and an exploration part, which in the form of regret penalizes the most often selected channels.

In **cooperative environments**, different secondary users share their sensing outcome to improve the detection performance at the expense of increased latency and communication overhead. The essential optimizations regarding the spectrum sensing in cooperative environments, consider the aspects of node selection, channel selection and node clustering. The node selection and node clustering strategies [32, 33] are similar in the sense that they aim to select a subset of the secondary users based on the nodes' characteristics in order to reduce the sensing latency and control traffic overhead. The channel selection strategies essentially adopt the single user sensing strategies (in terms of the sensing policies, exploration-exploitation trade-off, etc.) and extend them to the multi-user scenarios.

4.4.2 Spectrum Sharing Strategies

The spectrum sharing functionality exploits the radio context information for efficient secondary allocation, access, sharing and utilization of the spectrum opportunities, maintaining the secondary system QoS while providing primary system protection. Thus, in sensing-enabled C-MAC protocols, the performance of the spectrum sharing depends on the reliability of the outcome of the spectrum sensing functionality. The spectrum sharing is tightly related with the control channel management, since all of the required radio context information and

Fig. 4.2 Block diagram of
the spectrum sharing
processes
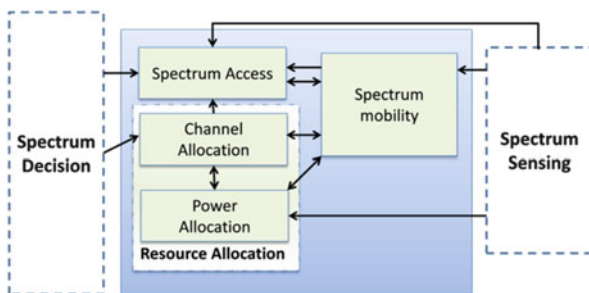
spectrum sharing/access decisions should be communicated between the different
CRN entities via reliable and secure control channel.

The spectrum sharing techniques are usually classified as vertical and horizontal
techniques. The **vertical** techniques refer to opportunistic spectrum access and/or
spectrum mobility where the secondary users share the same spectrum band with
the primary users, by exploiting the features of *interweave*, *underlay* or *overlay*
spectrum sharing. The **horizontal** techniques concern with the inter/intra-network
spectrum sharing between the secondary users. Figure 4.2 schematically depicts
the main processes related to the spectrum sharing functionality. These generic
processes are: resource allocation, spectrum access and spectrum mobility which is
of crucial importance when addressing the primary system protection requirement.

The following subsections aim to provide brief description of the most important
aspects with respect to the generic spectrum sharing processes.

4.4.2.1 Resource Allocation

The resource allocation comprises two generic processes, the ***channel allocation***
and the ***power allocation*** (Fig. 4.2), which are usually jointly optimized. The
channel allocation process is responsible for finding the most suitable frequency
and channel bandwidth, whereas the power allocation process is responsible for
managing the transmit power of the secondary users in order to satisfy the
interference constraints of the primary system. Thus, the resource allocation process
relates to resource parameters and resource allocation constraints. The *resource
parameters* define the space of optimization of the resource allocation process.
There are several types of resource parameters such as spectrum related parameters
(i.e. frequency, bandwidth, etc.), physical layer parameters (modulation, coding,
power, antenna configuration, etc.) and higher layer parameters (scheduling, buffer
management, ARQ, etc.). The set of *resource allocation constraints* is in general,
scenario specific and depend on the underlying CR use-case. However, they always
relate to either of the two generic C-MAC functional requirements: primary system
protection and maintaining the secondary system QoS.

Regarding the optimization the problem of resource allocation is multi-objective. The latest advances in the area of resource allocation optimization focus on two distinct methodologies: the classical methods and metaheuristcs based optimization algorithms. *Classical optimization methods*, like *single-variable* and *multivariable optimization*, convert the multi-objective optimization problems in a single-objective problem using a preference-based strategy and therefore they are usually inappropriate for solving the resource allocation problem. Oppositely, the *metaheuristcs* based algorithms utilize a population of solutions instead of focusing on a single one which makes them very suitable for solving multi-objective optimization problems. The most commonly utilized metaheuristcs based algorithms in terms of the resource allocation process vary from *ant colony optimization* [34] and *Swarm intelligence algorithms* [35] up to *Genetic algorithms* [36], *Differential evolution* [37] and *Simulated annealing algorithms* [38].

4.4.2.2 Spectrum Access

The spectrum access enables both the vertical and horizontal spectrum sharing by using CRN suitable multiple access protocols and is tightly related to the spectrum decision and channel allocation processes as well as the spectrum sensing functionality outcome. The multiple access techniques in CRN, rely on and exploit the features of the common multiple access protocols (extensively used in legacy wireless systems) such as Carrier Sense Multiple Access (CSMA) [39], Orthogonal Frequency Division Multiple Access (OFDMA) [40], Space Division Multiple Access (SDMA) [41], Ultra-Wideband/Code Division Multiple Access (UWB/CDMA) [42], Time Division Multiple Access/Frequency Division Multiple Access (TDMA/FDMA) [43] and Dynamic Frequency Hopping (DFH) [44]. Table 4.3 summarizes the most important aspects of multiple access techniques in the context of CRN.

The spectrum access can be either autonomous or coordinated. The ***autonomous spectrum access*** is accomplished by achieving individual goals like the QoS requirements [27, 45] or the energy consumption [46] of a given secondary user. Additionally, it may employ learning techniques based on past decisions and outcome, which ultimately can increase the system throughput [47] and achieve autonomous load balancing [20]. The ***coordinated spectrum access*** is more efficient in terms of the achievable CRN performance and requires more complex C-MAC protocols. As an example, in centralized scenarios the C-MAC can manage the process of sharing the primary system environmental knowledge, which can yield increased spectrum awareness [27], but will inevitably increase the implementation complexity.

Table 4.3 Multiple access schemes comparison

	CSMA	OFDMA	SDMA	UWB/CDMA	TDMA/FDMA	DFH
Sharing strategy	Interweave/underlay	Interweave/underlay	Underlay/overlay	Underlay	Interweave/underlay	Interweave
Multiple access dimension	Time	Frequency	Space	Spread spectrum	Time/frequency	Time/frequency
PU protection	Collision avoidance/interference threshold	Collision avoidance/interference threshold	Interference mitigation/interference alignment/cooperation	Interference threshold	Collision avoidance/interference threshold	Collision avoidance
Requirement from the spectrum sensing	Available frequency bands/available power mask	Available frequency bands/available power mask	Fading gains between SUs and PUs/available power mask	Available power mask	Available time slots/frequency bands/available power mask	Available frequency bands
Implementation complexity	Low/moderate	Moderate	High	High	Moderate	low
Computation complexity	Low/moderate	Moderate/high	High	High	Moderate	low

4.4.2.3 Spectrum Mobility

The spectrum mobility is a direct facilitator of the primary system transparency requirements and enabler of the concept of frequency agility which results in increased secondary system performances. The spectrum mobility can be realized through four types of spectrum handovers: static, reactive, proactive and hybrid [48].

In the case of *static* spectrum the secondary system will stay on the same channel and not transmit until the same channel becomes free again. The biggest setback of this spectrum handover type is the high secondary system latency incurred by the transmission of the primary system. In *reactive* spectrum handovers, the secondary user vacates the licensed channels after the reappearance of a primary user. The efficiency of the reactive spectrum handovers is tightly related to the handover latency. The *proactive* spectrum handovers utilize some predictive methods that trigger when the secondary users must vacate the underlying channel which minimizes both, the handover latency and the number of future spectrum handovers. These types of spectrum handovers can potentially benefit from learning and prediction, where the CRN can learn the environment dynamics, predict undesirable situations and act to avoid such situations in timely manner [49]. However, employing learning and prediction techniques increase the computational complexity. Finally, the *hybrid* spectrum handovers represent a compromise between the high latency (reactive spectrum handovers) and high complexity (proactive spectrum handovers). Typical representative of the hybrid approaches is the Incumbent Detection Recovery Protocol (IDRP) used in IEEE 802.22 WRAN standard [2].

4.4.3 Control Channel Management Strategies

The control channel (CC) is used for exchange and dissemination of different types of sensing and sharing related data (such as spectrum sensing outcome, sharing decisions, channel access feedback, etc.) between the different entities in the CRN. Thus, the CC provides its services to number of CRN operational aspects such as network self-organization, network coordination, synchronization, cooperation, collaboration, spectrum mobility, flexible data connections and increasing and attaining overall spectrum efficiency [50]. However, due to spectrum heterogeneity, the CC management functionality in C-MAC, encounters several major issues, not present in the legacy C-MAC protocols. CC variable availability, saturation, coverage, security are among the most common CC design challenges that require addressing. Depending on the target application and operational mode of the CRN, the CC management functionality of the C-MAC protocol should decide how to establish the CC to mitigate the mentioned problems [51]. This section briefly covers several popular techniques for CC establishment in spectrally heterogeneous environments.

The CC can be established as either Dedicated or Non-dedicated. The **Dedicated CC (DCC)** solutions refer to CC establishments where a set of secondary spectrum resources are solely dedicated for the transport of secondary system control information. The DCC is usually realized as a Global, Local or Dynamic. When establishing *Global DCC (GDCC)* [31], set of primary user channels are globally allocated for the CC and the secondary users tune to these channels to exchange control information. The *Local DCC (LDCC)* is similar to the GDCC except that it is established locally, on secondary users group [52], or a secondary user cluster level [53]. The LDCC alleviates the single point of failure effect exhibited by the GDCC. When both, the GDCC and the LDCC establishments adapt to the varying primary network conditions, the *Dynamic DCC (DDCC)* is established [54]. Although, the DDCC provides largest secondary system performance gain, it is also the most complex DCC establishment.

The other group of technical solutions for CC establishment consists of **Non-Dedicated CC** (NDCC) proposals. In the NDCC enabled CRN, the secondary users share the available channels of the target licensed band for exchange of both control and data packets. The NDCC can be established by Frequency Hopping, Rendezvous or as Ultra-Wideband. When establishing *Frequency Hopping NDCC (FHNDCC)* [55], the secondary users hop across the available channels to exchange control information, using a predetermined hoping list. However, this solution requires tight synchronization. When using *Rendezvous NDCC (RNDCC)* [56], the secondary users hop across the channels using different hoping lists that overlap multiple times. Thus, the RNDCC can be either synchronous or asynchronous. One of the main parameters of interest in RNDCC establishments is the Time-to-Rendezvous (TTR) denoting the average time that elapses until the first overlap occurs. The hopping lists should be designed to minimize the TTR while providing as many overlaps as possible. The last CC establishment is via Ultra-Wideband technology. The *Ultra-Wideband NDCC (UWBNDCC)* [57] is established in underlay fashion, by spreading the control information over the whole disposable band. However, this establishments suffers from all common drawbacks of UWB technology such as limited transmission range, interference to the primary system etc. Table 4.4 summarizes the discussed characteristics and behaviours of the various CC establishments.

The following section presents three particular C-MAC algorithms. Each of them fit in specific generic area (i.e. spectrum sensing, spectrum sharing and control channel management) and implement different specific functionalities, placing them in the generic C-MAC cycle layout. Besides emphasizing the specifics of the C-MAC protocol design process for a particular operational scenario, the use cases also serve as an illustration of the potential benefits of applying the C-MAC cycle on certain C-MAC protocols in terms of understanding the range of applicability of the protocols and their basic operational limitations.

Table 4.4 Qualitative comparison of various control channel establishments techniques

	Dedicated control channel			Non-dedicated control channel		
	Global	Local	Dynamic	Frequency hopping	Rendezvous	Ultra wideband
Establishment	Global	Local	Dynamic	Frequency hopping	Rendezvous	Ultra wideband
Saturation susceptibility	High	Medium	Low	Medium-low	Low	High
PU activity robustness	Low	Medium	High	High	High	High
Coverage	Global	Local	Global/local	Link-based	Link-based	Global/local
Security attacks resilience	Low	Medium-low	High	High	High	Medium-low
Configuration delay	Low	Medium-high	Medium-high	Medium	Low-medium	Medium
Access delay	Low	Low	Low	High	High-medium	Low

4.5 C-MAC Cycle Use Cases

This section presents three chosen C-MAC cycle use cases, aiming to elaborate the implemented functionalities in each of them. They were chosen to illustrate particular generic C-MAC features (elaborated in Sect. 4.4). In particular, Sect. 4.5.1 presents the cooperative spectrum sensing based on Estimated Noise Power as a representative of the spectrum sensing Sect. 4.5.2 introduces a horizontal spectrum sharing strategy for secondary systems based on Node Clustering and Beamforming, as a representative of a spectrum sharing. Finally, Sect. 4.5.3 analyzes, assesses and explains the modularly merged within the C-MAC cycle of the multiuser quorum-based multiple rendezvous strategy for control channel establishment in distributed Cognitive Radio Networks. These C-MAC cycle use cases are selected in a relation to the expertise of the authors in the area of C-MAC protocol engineering and they are result of the authors' recent work in this area.

4.5.1 Cooperative Spectrum Sensing Based on Estimated Noise Power

Spectrum sensing has been pinpointed as one of the facilitating technologies for spectrum opportunities detection in Dynamic Spectrum Access (DSA) and CR systems. Although the latest developments in CR regulation favor the database approach [1], spectrum sensing can provide reliable opportunistic access in many scenarios (e.g. dynamic and unpredictable environments [58]). Under these circumstances, the efficiency and reliability of the spectrum sensing approach can prove to be crucial in providing the required system performance to the underlying cognitive radio network. More accurate and efficient spectrum sensing results in higher spectral utilization of the secondary user system as well as in lower interference to the incumbent system. The Cooperative Spectrum Sensing (CSS) approach increases the sensing performance by exploiting the spatial diversity from multiple secondary user sensing nodes.

This subsection focuses on a C-MAC cycle use case regarding the Cooperative Spectrum Sensing process. More specifically it elaborates on a specific Cooperative Spectrum Sensing approach which utilizes the aspects of Noise Power Estimation (ENP). Figure 4.3 depicts the C-MAC instantiation of the *Cooperative Spectrum Sensing based on ENP*. This approach is designed to perform single band spectrum sensing and exploit the sensed information from multiple CR users. The CSS based on ENP approach heavily depends of the cooperative/collaborative aspects of the C-MAC cycle and tightly relates to the synchronization feature (as a facilitator of the cooperation and collaboration process). The remainder of this subsection will elaborate and discuss in more details the CSS based on ENP and its relation to the C-MAC protocol.
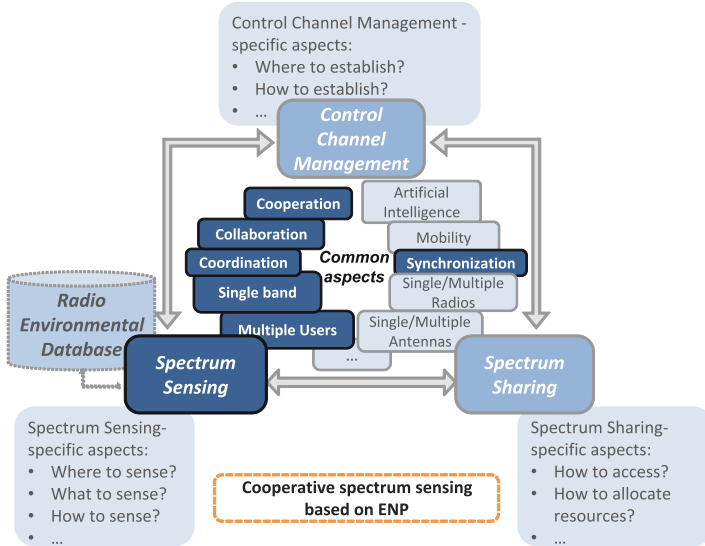
Fig. 4.3 CSS based on ENP C-MAC cycle instantiation

4.5.1.1 The ENP Theory

Regarding the spectrum sensing process, the simplest method capable of detecting the presence of a primary user is based on the energy detection notion. The Energy Detector (ED) is most optimal when there is no information about the primary user (i.e. no feature detection can be performed). Likewise, the ED is appropriate for fast and wideband spectrum sensing where different kinds of signals exist. The ED performance has been studied in many works where the receiver has perfect knowledge about the noise power. However, in practical receivers (spectrum sensors) it is impossible to have perfect knowledge about the noise power level. A set of works have addressed this problem [59, 60] stating that there exists a minimum value of SNR under which the detection is not possible even for infinite number of sensed samples. This phenomenon is known as the SNR wall. It has been preserved as an inexorable problem, since the estimation of the noise power always differs from the real noise power [61]. Recent studies [62, 63] have proved that the SNR wall is not caused by the presence of an uncertainty in the noise power itself, but by insufficient information about the noise power estimation. The authors in [62] propose an approach that is capable of alleviating the SNR wall by utilizing higher number of noise-only samples, called the *Estimated Noise Power (ENP)*. The work in this subsection extends the given theory by introducing the aspects of Equal Gain Combining (EGC) and Majority Voting (MV) cooperative spectrum sensing to the ENP approach.

4.5.1.2 System Model and Analytical Relations

This section briefly describes the system model used for deriving the analysis of the proposed CSS approach and derives analytical forms for the probability of detection and probability of false alarm for the EGC and MV techniques. The system model assumes signal detection in an AWGN channel where the ith received signal sample under both hypotheses is given as:

$$y_i = \begin{cases} n_i, & H_0 \\ x_i + n_i, & H_1 \end{cases} \tag{4.1}$$

where H_0 and H_1 denote the signal's absence and signal's presence hypothesis, respectively. The signal sample x_i and noise sample n_i are circularly-symmetric complex Gaussian random variables with zero mean and variances: $2S$ and $2\sigma^2$, respectively. Based on Eq. (4.1) and assuming an energy detector, the detection problem can be defined as a log likelihood ratio test:

$$\Lambda_g(y) = \frac{1}{2\sigma^2} \frac{1}{N} \sum_{i=0}^{N-1} |y_i|^2 \begin{matrix} > \xi & H_1 \\ < \xi & H_0 \end{matrix} \tag{4.2}$$

where N denotes the number of sensed samples and represents the decision threshold. For a predefined value of the probability of false alarm (P_{fa}) and number of sensed samples (N), the detector must know the noise variance in order to set its decision threshold ξ. Because in practical implementations the energy detector utilizes only an estimate, instead of the true noise power (i.e. variance) the likelihood ratio test is defined as the generalized likelihood ratio test when no a priori knowledge about the primary signal is available:

$$\Lambda_g(y) = \frac{1}{2\hat{\sigma}^2} \frac{1}{N} \sum_{i=0}^{N-1} |y_i|^2 \begin{matrix} > \xi & H_1 \\ < \xi & H_0 \end{matrix} \tag{4.3}$$

where $\hat{\sigma}^2$ denotes the noise power estimates and depends only on the specific estimation technique in use. In the proposed system model, the noise power samples at the receiver are estimated utilizing the ML approach:

$$\hat{\sigma}_{ML}^2 = \frac{1}{2M} \sum_{i=1}^{M} |n_i|^2 \tag{4.4}$$

where n_i denotes the noise samples, while M refers to the number of sensed noise samples. As discussed in [62, 63], the $\hat{\sigma}_{ML}^2$ estimator represents an effective noise estimator capable of reaching the Cramer-Rao bound.

Based on these system model assumptions and as elaborated in [63], the probability of detection ($Q_{d\_ENP}^{EGC}$) and probability of false alarm ($Q_{fa\_ENP}^{EGC}$) for the

EGC based ENP approach are defined as:

$$Q_{d\_ENP}^{EGC} = Q \left(\frac{\frac{\xi}{1+SNR} - 1}{\sqrt{\frac{M+N}{KMN}}} \right) \tag{4.5}$$

$$Q_{fa\_ENP}^{EGC} = Q \left(\frac{\xi - 1}{\sqrt{\frac{M+N}{KMN}}} \right) \tag{4.6}$$

where $Q(\cdot)$ denotes the Q-function, while SNR and K represent the received SNR at the CR node and the number of cooperating nodes respectively. Similarly the probability of detection ($Q_{d\_ENP}^{MV}$) and probability of false alarm ($Q_{fa\_ENP}^{MV}$) for the MV based ENP approach are defined as:

$$Q_{d\_ENP}^{MV} = \sum_{i=K/2}^{K} Q \left(\frac{\frac{\xi}{1+SNR} - 1}{\sqrt{\frac{M+N}{MN}}} \right)^{i} Q \left(\frac{\frac{\xi}{1+SNR} - 1}{\sqrt{\frac{M+N}{MN}}} \right)^{K-i} \tag{4.7}$$

$$Q_{fa\_ENP}^{MV} = \sum_{i=K/2}^{K} Q \left(\frac{\xi - 1}{\sqrt{\frac{M+N}{MN}}} \right)^{i} Q \left(\frac{\xi - 1}{\sqrt{\frac{M+N}{MN}}} \right)^{K-i} \tag{4.8}$$

where $Q(\cdot)$ denotes the Q-function, while SNR and K represent the received SNR at the CR node and the number of cooperating nodes respectively. From Eqs. (4.5) to (4.8) it is evident that the CSS based on ENP represents a complex process which exploits a set of *different sensing parameters* (e.g. number of sensed samples, number of noise samples, number of cooperating users, threshold adaptation). This implies that the C-MAC protocol must perform an "intelligent" selection of the given sensing parameters in order to achieve the required detection performance. The following subsection elaborates on the performance analysis (i.e. numerical and experimental validation) of the proposed CSS based on ENP approach.

4.5.1.3 Performance Analysis

This section briefly discusses the performance of the EGC and MV cooperative spectrum sensing techniques when utilizing the ENP theory and validates the theoretical assumptions with experimental results. The analysis is performed in terms of the ROC curve, average Bayesian risk and detection probability. Moreover, the values of the input parameters (i.e. K, N, M, etc.) in the analysis are selected based on practical investigations and on the studies made in previous works

Fig. 4.4 ROC curve
($N = 10^6$, $M = 10N$,
$K = 10$,
$SNR = 5 * 10^{-4}$) [64]
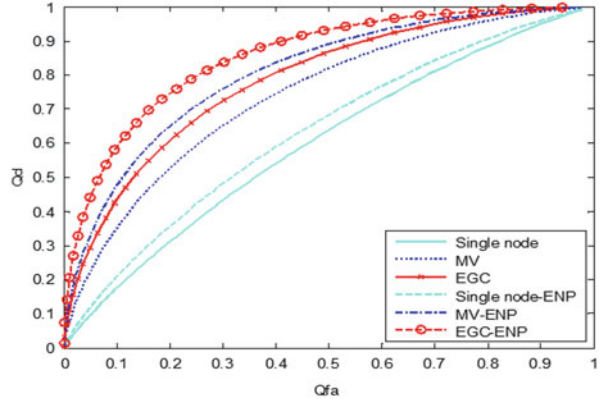
regarding the ENP approach [62,63]. In order to simplify the notation, the MV based ENP and EGC based ENP techniques will be noted as MV_ENP and EGC_ENP, respectively.

Figure 4.4 delineates the ROC curves for single node detection and for both EGC and MV techniques. As seen from the figure, the non-cooperative case (without the ENP approach) achieves the worst, while the EGC_ENP achieves the best ROC performance. It is also evident that MV_ENP outperforms the classical EGC fusion rule due to the utilization of the ENP approach.

Figure 4.5 depicts the average Bayesian risk in terms of the number of sensed signal samples N. The average Bayesian risk defines the overall performance of the cooperative spectrum sensing and can be expressed as follows:

$$\overline{R} = P(H_0)Q_{fa}C_{10} + P(H_1)(1 - Q_d)C_{01} \qquad (4.9)$$

where $P(H_0)$ and $P(H_1)$ represent the probability of primary user absence and presence, respectively, while C_{10} and C_{01} denote the price coefficients when making a wrong decision (in this work $C_{10} = C_{01} = 1$) and are used to model the system behavior and type [63].

It is evident from Fig. 4.5 that EGC_ENP achieves the best performance, i.e., the smallest Bayesian risk, while the non-cooperative case without ENP incorporates the highest Bayesian risk, i.e., worst performance. Similarly to the conclusions from the previous figure, MV_ENP outperforms the classical EGC and achieves lower Bayesian risk.

In order to validate the CSS based ENP theory, the analytical models of the EGC_ENP and MV_ENP are compared to experimental results regarding the probability of detection. Figure 4.6 depicts the theoretical vs. experimental results regarding the EGC_ENP approach and proves the validity of derived EGC_ENP analytical models. Moreover, it can be noticed that the EGC_ENP has a constant 3 dB detection performance increase with each sample size quadrupling.

Fig. 4.5 Average Bayesian risk in dependence of the number of sensed samples N ($M = 10N$, $K = 10$, $Q_{fa} = 0.1$, $SNR = 5 * 10^{-4}$, $P(H_0) = 0.5$) [64]
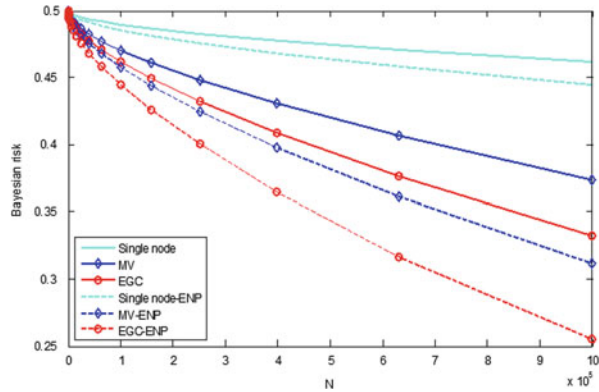
Fig. 4.6 Q_d vs. SNR theoretical and experimental EGC_ENP result ($K = 5$, $M = 10^7$, $Q_{fa} = 1\%$)
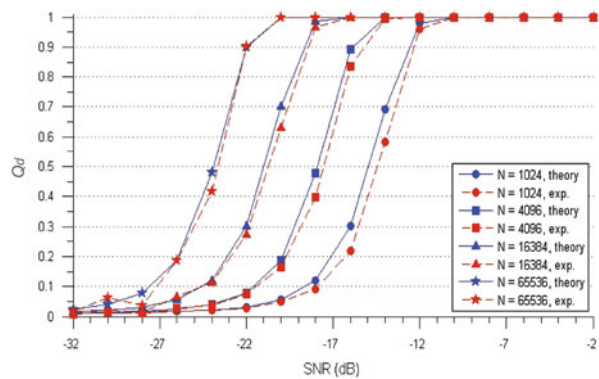
Fig. 4.7 Q_d vs. SNR theoretical and experimental MV_ENP result ($K = 5$, $M = 10^7$, $Q_{fa} = 1\%$)
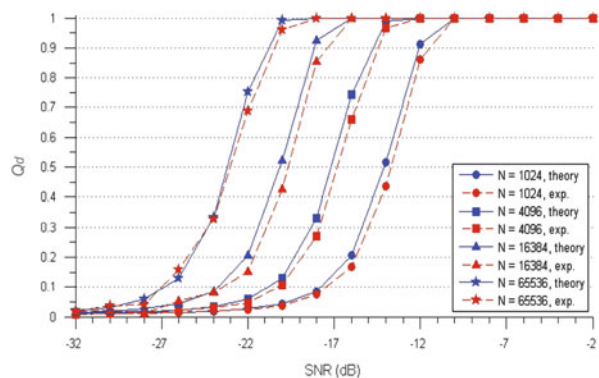
Figure 4.7 depicts the theoretical vs. experimental results regarding the MV_ENP approach and proves the validity of derived MV_ENP analytical models. Similarly to the conclusions from the previous figure, the MV_ENP achieves a constant 3 dB detection performance increase with each sample size quadrupling.

The subsection presented a specific C-MAC cycle use case regarding the spectrum sensing process which exploits the features of CSS and introduces the aspect of ENP. As discussed, the CSS based on ENP represents a multifaceted process which utilizes a variety sensing parameters that need to be governed in the most efficient manner from the C-MAC in order to achieve the required performance. This C-MAC related task tightly relates to the requirements and goals of the underlying use-case scenario. For example, in many cases the best sensing setup is optimal for achieving the best signal detection performance, however can be suboptimal regarding other system related performances like, throughput, packet delay, etc. Therefore, it is more efficient for the C-MAC to coordinate and optimize the sensing process regarding the system parameters of interest, rather than the signal detection. In this case the C-MAC related optimization can be based on the derived and validated analytical models for the probability of detection and probability of false alarm of the EGC_ENP and MV_ENP approaches.

4.5.2 Coordinated Beamforming for Spectrum Sharing

To guarantee high spectrum efficiency while mitigating the interference to the primary users, the CR should be able to adapt to the spectrum conditions flexibly. However, the interference caused by horizontal spectrum sharing becomes an obstacle that limits the system performance, such as the system throughput. Thus, in horizontal sharing, the goal is to try to find a way to decrease the inter-system interference (i.e. collisions) and increase its system throughput. The cooperation between the secondary systems can additionally increase the spectrum efficiency of the horizontal sharing process. One possible approach that is based on cooperation and can provide high spectrum utilization is the Space-Division Multiple Access (SDMA) approach.

This subsection elaborates on a specific C-MAC cycle use case regarding the horizontal spectrum sharing process. The given use case, entitled *Network Coordinated Beamforming with user Clustering (NCBC)* [65], utilizes the concept of SDMA via Network Coordinated Beamforming (NCBF) and user Clustering. Figure 4.8 depicts the C-MAC instantiation of the NCBC. This approach is envisioned to perform single band spectrum sharing and exploit the spatial diversity from multiple antennas and CR users. The NCBC depends of the cooperative/collaborative aspects of the C-MAC cycle and tightly relates to the distribution features of the C-MAC protocol in order to achieve the required system performance. The remainder of the subsection will discuss in more details the NCBC and its relation to the C-MAC protocol.

4.5.2.1 System Model

The system model considers the case where a set of K secondary CR (secondary user) systems coexist and share the same spectrum (frequency band). It is assumed
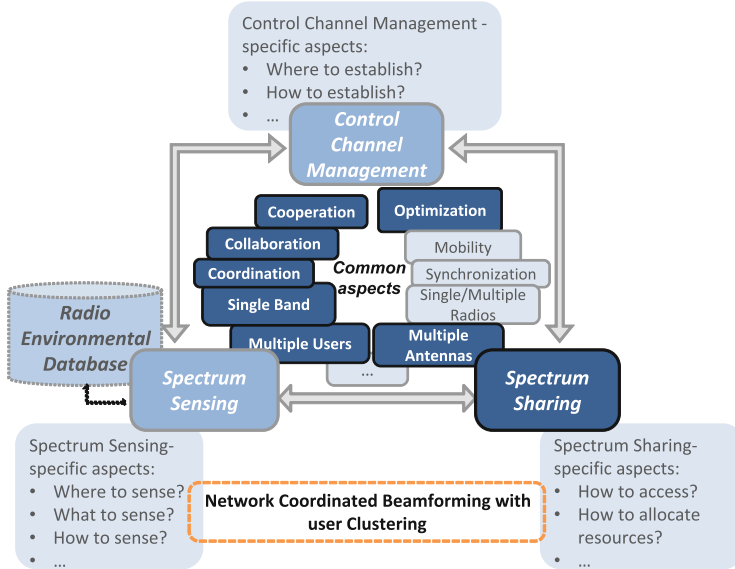
Fig. 4.8 SDMA based SU-SU sharing C-MAC cycle instantiation

that their coverage area overlaps totally or in some parts. Additionally, every system has one Base Station (BS) and multiple users. The BS serves only one user at a time, and interferes with remaining secondary systems, as shown on Fig. 4.9. This definition of the system model can be mapped onto an indoor scenario where multiple secondary systems, spatially collocated, use the same spectrum availability. An example use-case can be the event where a given TV White Space (TWS) is opportunistically shared by multiple LTE femto cells and IEEE 802.11af Access Points (AP), located in the same object.

Due to the overlapping of the systems, the served users will encounter large inter-system interference. One possible solution for mitigating the interference, thus enabling efficient spectrum sharing, is the SDMA approach, for example the NCBF solution. The following subsection will introduce the NCBC approach as a novel spectrum sharing method and will focus on its features and performance.

4.5.2.2 Network Coordinate Beamforming with User Clustering

The NCBF considers a multi-user MIMO channel where all BSs are equipped with one transmit antenna and all users have N_r receive antennas. As already elaborated, only one user per system is served at a given time, hence the total number of served users at any given moment is equal to number of coexisting systems K. It is assumed that all BSs cooperate ideally to compute the transmit beamforming and receive combining vectors [65]. The channel between all BSs and the kth user is represented
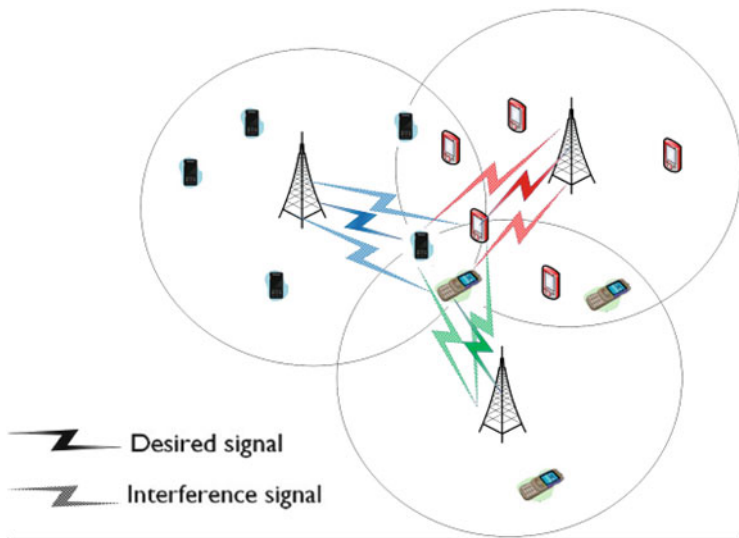
Fig. 4.9 System model example
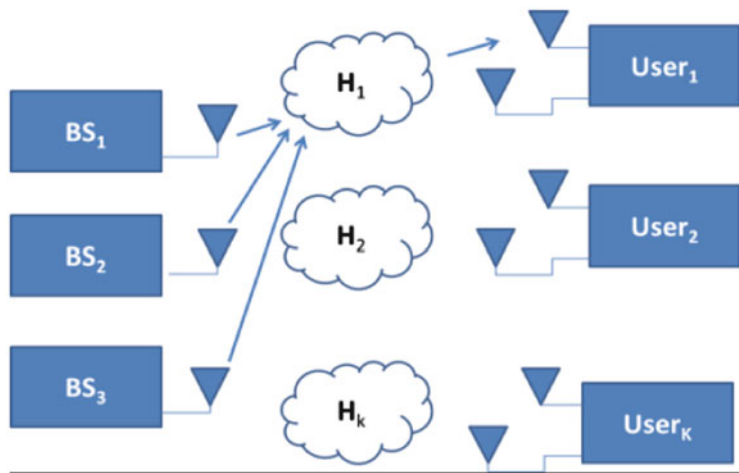
Fig. 4.10 Block diagram of NCBF

by the channel matrix $\mathbf{H}_k$ of size $N_r * K$ and has complex entries for the channel gains. Thus, every column of $\mathbf{H}_k$ represents the single-output multiple-input (SIMO) channel between each BS and the kth user. Figure 4.10 depicts the block diagram example of the NCBF.

The system operates in TDD mode in which the temporal variations of the channels are slow compared to the duration of the data frame [66]. Let x_k and $\mathbf{n}_k$ be the transmit symbol and the noise vector with variance σ_k^2 of the kth user,

respectively. Let $\mathbf{m}_k$ denote the transmit beamformer and $\mathbf{w}_k$ denote the receive-combining vector at the kth user. The received signal at the kth user can be expressed as:

$$y_k = \mathbf{w}_k^H \mathbf{H}_k \mathbf{m}_k \sqrt{p_k} x_k + \mathbf{w}_k^H \mathbf{H}_k \sum_{l=1, l \neq k}^{K} \mathbf{m}_l \sqrt{p_l} x_l + \mathbf{w}_k^H \mathbf{n}_k \qquad (4.10)$$

where p_k is the allocated transmit power of the BS for the kth user. In the case of coordinated beamforming strategies, the transmitter chooses $\mathbf{m}_k$ such that the subspace spanned by its columns lies in the null space of $\mathbf{w}_k^H \mathbf{H}_k (\forall l \neq k)$, i.e. $\mathbf{w}_k^H \mathbf{H}_k \mathbf{m}_k = 0 (\forall l \neq k)$ and $k = \overline{1, K}$. The authors in [65] have proposed two NCBF algorithms (linear and non-linear approach) that compute the beamforming and receive combining coefficients, which perfectly mitigate the inter-system interference. However, when the number of active users per system is larger than one, the computational complexity of the NCBF increases dramatically making the approach not suitable for real time operation.

The NCBC method decreases the complexity of NCBF by serving a group (i.e. cluster) of users in the given system with the same beamforming coefficients based on the correlation of their channel matrices. For this purpose, every BS has to calculate the correlation of the channel matrices between all of its active users. If the correlation between them is larger than a predefined threshold, they can be served with the same beamforming and receive combining coefficients. This approach can substantially decrease the computational complexity of the legacy NCBF. The correlation between two users can be computed as [65]:

$$\gamma_{corr} = 1 - \|\lambda_k - \lambda_l\|_F \qquad (4.11)$$

where λ_k denotes the eigenvalue matrix of the kth user's channel matrix $\mathbf{H}_k$, λ_l denotes the eigenvalue matrix of the kth user's channel matrix $\mathbf{H}_l$ and $\|\cdot\|_F$ denotes the Frobenius norm. If $\gamma_{corr} = 1$, the users will be completely correlated, while for $\gamma_{corr} = 0$, the users will be completely decorrelated. Lower will result in lower system performance and larger inter-system interference. The correlation threshold depends on the required user performance. For example, if the users require high data rate service, than the correlation threshold has to be closer to its maximum value (i.e. $\gamma_{corr} \approx 1$). In case when low delay latency is required (more agile system performance i.e. faster calculation of the beamformers), and the data rate is not crucial, the correlation threshold can attain lower values. The following subsection elaborates on the performance analysis of the proposed NCBC approach.

4.5.2.3 Performance Analysis

This section gives a brief insight into the performance of NCBC in terms of Signal to Interference Ratio (SIR) and the aggregate system capacity. Additionally, it

Table 4.5 Channel model
parameters [69]

Channel parameters	
Channel type	NLOS indoor
Delay spread	<100 ns
Number of Paths	8
Bandwidth	10 MHz
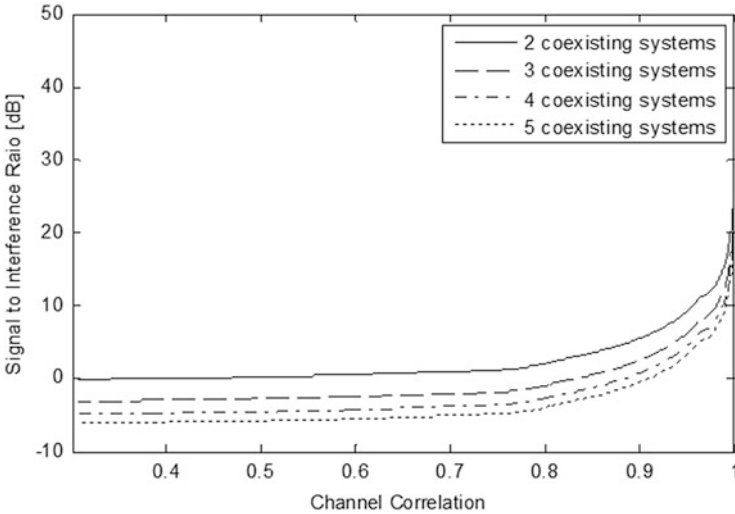
Fig. 4.11 Signal to interference ratio in dependence of the channel correlation

compares the performance of NCBC with a common Frequency Division Multiple
Access Spectrum Sharing (FDSS) technique that is based on equal frequency
division between all coexisting secondary CR systems. To obtain relevant results,
Monte Carlo simulations are carried out for all performance metrics. The channel
model used in the simulation analysis is based on the Kronecker MIMO channel
model for indoor environments [67] and its parameters are given in Table 4.5. For
simplification of the analysis, it is assumed that all BSs use equal allocation of the
transmit power for every user, hence denoting $p_k = p$.

The correlation between the channel matrices of the given two users plays a
crucial role of the system performance of NCBC. Figure 4.11 depicts the SIR in
dependence of γ_{corr} for different number of coexisting secondary CR systems. As
seen from the results, the SIR ratio decreases as γ_{corr} decreases, but even for low
channel correlation the SIR level does not fall below 5 dB. This SIR level can be
satisfactory for a set of different scenarios of modern day wireless systems [68],
meaning that in given circumstances it is possible for all users (of a given system) to
use the same beamforming and receive combining coefficients. This will drastically
decrease the computational complexity of the approach.
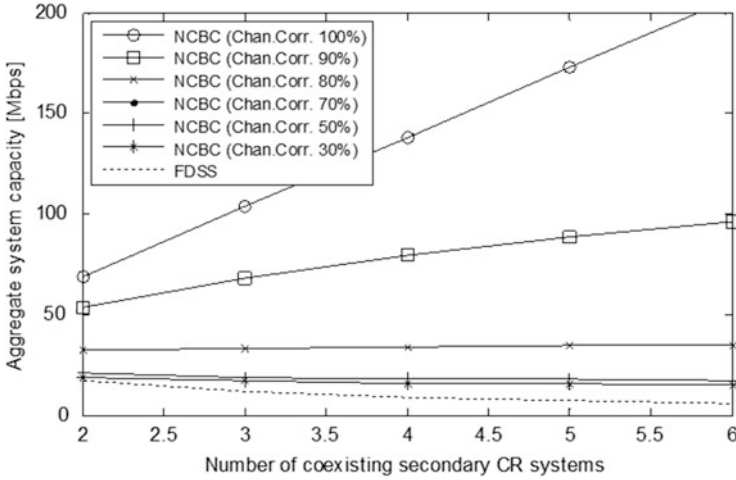
Fig. 4.12 Aggregate system capacity in dependence of the number of coexisting secondary CR systems ($SNR = 10\,\mathrm{dB}$)

The proposed NCBC method is also compared to a simple legacy FDSS method that divides the available spectrum to equal parts in bandwidth, between the coexisting secondary systems. In this manner, FDSS attains equal sharing, i.e. fairness between all coexisting systems.

Figure 4.12 depicts the aggregate system capacity in dependence on the number of coexisting secondary CR systems for both methods. NCBC outperforms FDSS for any case and the performance gain increases as the number of coexisting system increases. This is due to the fact that the spectral efficiency of any CBF scheme rises as the number of transmit and receive antennas increases. In the case of NCBC, the number of transmit antennas is mapped onto the number of BSs, i.e. number of coexisting systems K. For the FDSS scheme the system capacity decreases because the method splits the available spectrum into equal chunks of bandwidth for every coexisting system, thus resulting in decreased system capacity.

The subsection presented a particular spectrum sharing C-MAC cycle use case, denoted as NCBC, which utilizes the aspects of SDMA (i.e. NCFB) and user clustering. The NCBC represents a fundamental routine of the C-MAC protocol, which facilitates the *multiple access* scheme in CR systems. Regarding the common aspects and the operation of the C-MAC cycle, the C-MAC protocol is responsible for aiding the operation of NCBC by collecting and reliably distributing of the beamforming coefficients. Moreover, the C-MAC protocol aids the clustering i.e. grouping of the CR nodes, which is tightly related to the channel correlation coefficient i.e. the required user and system performance (i.e. QoS demands, real-time system operation, etc.), which were previously discussed.
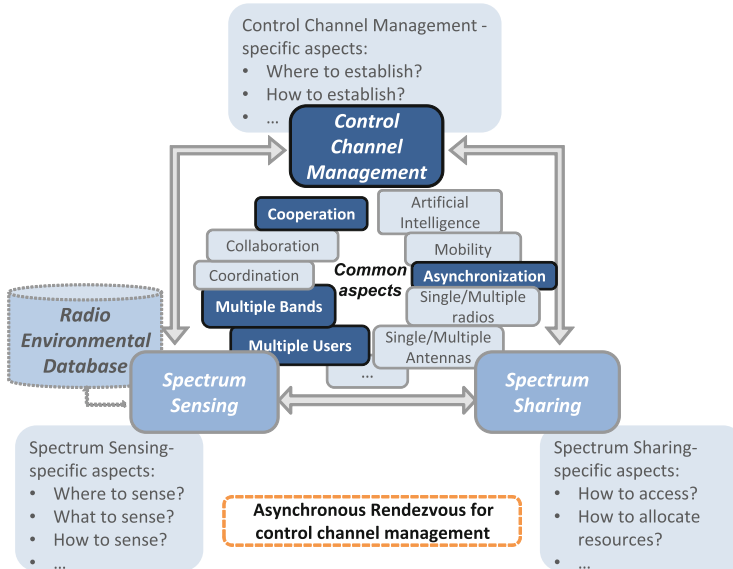
Fig. 4.13 Asynchronous rendezvous CC management C-MAC cycle instantiation

4.5.3 Asynchronous Rendezvous for Control Channel Management

The DCC in multichannel CRNs facilitates the coordination, cooperation and collaboration between the spectrum sensing and the spectrum sharing of the involved CR entities. In particular, the DCC enables the neighbor discovery and control signaling in terms of exchange of spectrum measurement results, access/sharing decisions and feedback etc. However, the existence of dedicated channel for control data exchange in CRNs, especially in distributed environments, may not be always feasible. The DCC needs to be established on a vacant legacy channel accessible by the majority of CR nodes and not interrupted over a longer time period. Aiming to fulfill these requirements the CRs might encounter problems such as spectrum heterogeneity, saturation and single point of failure. In distributed environments, the asynchronous operation of the CR nodes might raise additional reliability and sustainability concerns.

This subsection targets the dynamic control channels, and in particular focuses on a rendezvous protocol for dynamic and asynchronous CC establishment, pinpointed as the most reliable and efficient in terms of the CRs operation, and most harmless and transparent for the primary user operation [51]. Figure 4.13 illustrates the C-MAC cycle instantiation for the rendezvous specific CC management. Namely, the rendezvous protocols operate in multi-band and multi-user environments, and require cooperation between the operating CR nodes in the search of a common vacant legacy channel to rendezvous and exchange the control traffic. Furthermore,
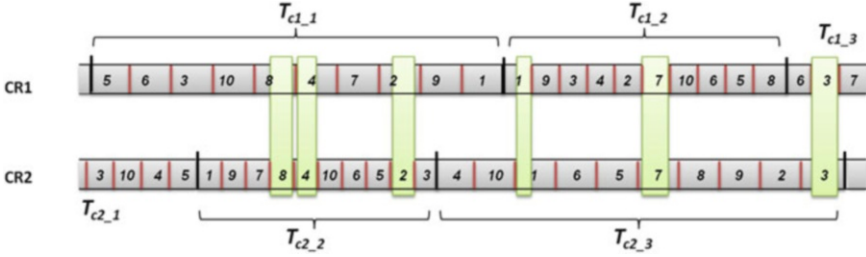
Fig. 4.14 The random rendezvous cycle duration and asynchronous operation provides overlapping between the both cognitive radios in the free channels (ch_i, $i = 1, \ldots, 10$)

the rendezvous control channels can enable, as well as exploit the asynchronous operation of the nodes to provide a faster rendezvous and control channel establishment.

The subsection focuses on a specific rendezvous protocol realization, i.e. the RAC^{2}E-gQS protocol for asynchronous rendezvous in cognitive radio ad-hoc networks. RAC^{2}E-gQS combines the asynchronous and randomness properties of the RAC^{2}E protocol and grid-quorum systems (gQS) to handle the channel heterogeneity and assure more rapid rendezvous. The remainder of the subsection is organized as follows. First, the details on the RAC^{2}E protocol operation are provided, followed by the explanation on the grid-quorum strategies for channel prioritization and mapping. At the end, the subsection provides the performance evaluation of the combined RAC^{2}E-gQS protocol and the concluding remarks.

4.5.3.1 RAC^{2}E Protocol

RAC^{2}E (Rendezvous protocol for Asynchronous Cognitive radios in Cooperative environments [70]) is a rendezvous protocol for distributed cognitive radio network environments. The protocol relies on an asynchronous operation of the nodes, eliminating the need of synchronization establishment, which is an especially difficult task in distributed environments. Moreover, RAC^{2}E fosters even an additional randomization among the nodes to ensure rapid rendezvous on a particular temporary unused channel from the primary system.

The operation of the rendezvous phase of the RAC^{2}E is illustrated on Fig. 4.14. Each cognitive radio aiming to establish a control channel independently selects a random rendezvous cycle duration of $T_{c\_i\_j}$ (ith cognitive radio, jth rendezvous cycle). This time duration is selected randomly with a uniform distribution in the range $[T_{min}, T_{max}]$, where $T_{min} = T_c \Delta T/2$, $T_{max} = T_c \Delta T/2$ and T_c represents the mean rendezvous cycle duration, while $\Delta T = kT_c$ is the randomization interval and k represents the randomization coefficient. The chosen $T_{c\_i\_j}$ interval is further segmented into M time slots, with each slot (having a duration of $\tau_{i\_j} = T_{c\_i\_j}/N$) assigned to a particular channel unoccupied by the primary users. As illustrated

Fig. 4.15 Rendezvous at
channel i event
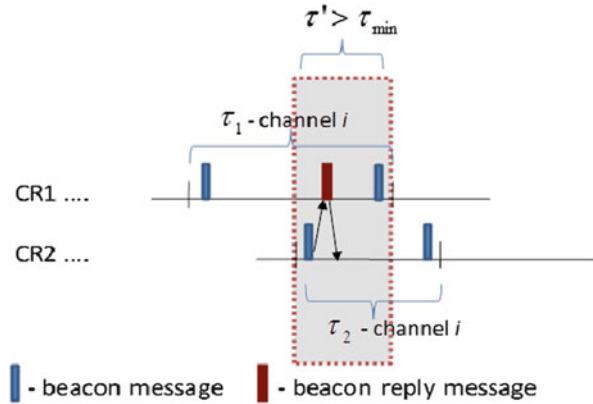
on Fig. 4.14, the randomization ensures that overlapping at same channels occurs randomly in wider or narrower time intervals.

The CR sends a short beacon message at the beginning and end of every slot τ_{i_j}, to signalize its presence in the channel. These particular times of beaconing are selected since they provide the highest probability of rendezvous between the CR nodes. In the meantime, between the both beacon messages, the rendezvous node aims to capture the beacons coming from the other CRs operating on the same channel. As Fig. 4.15 illustrates, the randomization (i.e. asynchronous operation of the both nodes) guarantees that at least one of the beacon messages will be delivered to the other nodes tuned to the same channel at the moment. This justifies the preference of a random $T_{c\_i\_j}$ duration (Fig. 4.14), which provides a more successful delivery of the beacon messages, in comparison with the synchronous case. A rendezvous occurs when two nodes are tuned to the same channel and they exchange at least one beacon and one beacon reply message. The condition $\tau > \tau_{min}$ must be fulfilled for the rendezvous to occur, where τ is the overlapping duration and τ_{min} is the required time for exchange and processing of both, the beacon and the beacon reply message (Fig. 4.15).

The mapping of channels into time slots in the rendezvous phase of the RAC^{2}E protocol is another important task. This can be done using several methods considering the channel priorities based on the channel ranking lists created in the sensing phase by each node independently. The combination of the RAC^{2}E protocol with the *grid-quorum based channel mapping* (Sect. 4.5.3.2) can yield a powerful **RAC^{2}E-gQS rendezvous protocol** for asynchronous operation in a distributed environment assuring rapid rendezvous between the cognitive nodes.

4.5.3.2 Grid Quorum Channel Mapping

Quorum-based algorithms [71, 72] recently became popular in the area of wireless networking, because of their capabilities to introduce resilience to node and

Table 4.6 4×4 grid: Pair-On-Pair (PoP): quorum (0,0) in bold

0	5	11	15
4	1	7	13
10	6	2	9
14	12	8	3

Table 4.7 4×4 grid: Diagonal (Diag): quorum (0,0) in bold

0	4	8	12
13	1	5	9
10	14	2	6
7	11	15	3

network failures. There are different types of QSs, within which a grid-based QS proposed by Maekawa [72], is widely utilized in power-saving (PS) protocols. In the grid-quorum systems, sites (elements) are logically organized in a grid in the shape of a square. There are two important properties that a grid-quorum system needs/should satisfy, i.e. the intersection property – the quorums need to intersect when perfectly synchronized, and a rotation closure property – the quorums need to intersect at least once regardless on the elements shifting.

The grid-based QS [56, 73], have recently become a research target in the area of rendezvous protocols, and specifically the channel mapping problems. There are two possible aspects of the grid-quorum systems in terms of the rendezvous protocols: the grid forming scheme and the channel-to-slot mapping scheme. Regarding the grid forming scheme the RAC^2E-gQS protocol considers two possible approaches [56, 73]: the Pair-on-Pair (PoP) grid forming and the Grid Diagonal (GD) grid forming, presented in Tables 4.6 and 4.7, respectively. Both schemes satisfy the intersection property, but only the GD scheme satisfies the rotation closure property.

Regarding the channel-to-slot mapping the RAC^2E-gQS also considers two approaches [56, 73]: a Row-Column mapping and a Diagonal mapping scheme (illustrated on Fig. 4.16 for the case of four candidate channels). The outcome of both algorithms provides an input to the channel hopping sequences called channel maps. Each node maps its channels according to the channel quality (based on the channel ranking lists formed in the sensing phase) without any exchange of information, where the better channels get priority. The period (cycle) of a channel map depends on the number of channels N and equals $M = N^2$ slots (selected from an $N x N$ grid). Both channel-to-slot mapping methods are designed for three or larger number of channels (i.e. $N > 3$).

Channels are mapped to grid indexes in both methods (channel 1 (C1) is mapped to index 0, channel 2 (C2) to index 1 etc.), where each channel in a CR network has its own index known by the nodes. A node adopts its map according to the priority of the channels, e.g., if a node A has the following map 2/4/3/1, channel C2 is the best, channel C4 is the second best etc. In the first method, the Row-Column mapping, in the first step (Step 1 in Fig. 4.16a), a node selects its channel map in a row-column manner, where the row number (channel number) is always equal to the column number (channel number). The best channel is channel 2, so it selects the quorum slots: 3, 6, 7, 12, 15 (Step 2), channel 3 is allocated to slots 0, 2 and
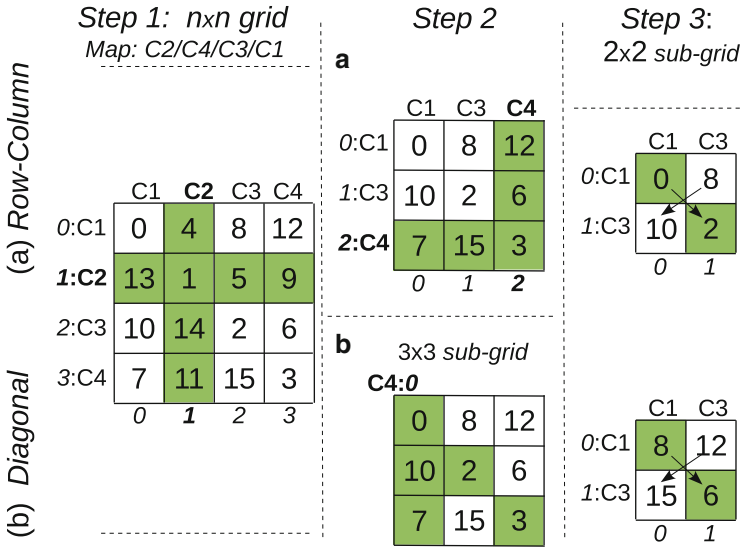
Fig. 4.16 Three steps of the channel-to-slot mapping: (**a**) Row-column, (**b**) Diagonal

channel 1 to slots 8, 10 (Step 3). The second method, Diagonal mapping, is similar to Row-Column mapping until a 3×3 sub-grid is obtained. The next channel is mapped (and sub-grid is cut accordingly) in a column-diagonal manner, selecting the first column and the main diagonal, e.g., channel 4 is mapped to slots 2, 3, 7, 10 (Step 2 in Fig. 4.16b). The last two channels are allocated as in the first method.

4.5.3.3   RAC²E-gQS Performance Analysis

This subsection demonstrates the performances of the RAC²E-gQS protocol for the different channel mapping methods elaborated above [56]. The simulation analyses envision a scenario with two cognitive radios aiming for a rendezvous on a certain common channel. Two cases are evaluated:

1. When both CRs have the same channel ranking lists, e.g. both have [1, 2, 3, 4, 5] as a priority list in the case of 5 candidate channels.
2. When the both CRs have completely different channel ranking list, e.g. CR1 has [1, 2, 3, 4, 5] while CR2 has [5, 4, 3, 2, 1] in the case of 5 candidate channels.

These two cases are considered since they provide the two extremes of rendezvous performances, i.e. they are the best and the worst case scenarios. One performance metric of interest in the analysis is the *average number of potential rendezvous (channel matchings)* per cycle which is in inverse proportion to the time-to-rendezvous (TTR). The second evaluated performance metric is the *inter rendezvous time variance*, representing the variance between two potential

Table 4.8 Minimum (Min), maximum (Max) and average (Mean) number of potential RDVs per cycle for gQS schemes in slot synchronized CRNs; No.c/s stands for Number of channels/slots; Ch. Rank. is the channel ranking lists

No.c/s	Ch. Rank.	Metric	PoP^{RC}	GD^{RC}	PoP^{DC}	GD^{DC}
5/25	Same	Min	1	1	3	3
5/25	Same	Mean	6.52	6.52	6.52	6.52
5/25	Same	Max	25	25	25	25
5/25	Different	Min	0	0	0	0
5/25	Different	Mean	3.56	3.56	3.56	3.56
5/25	Different	Max	7	7	7	7
10/100	Same	Min	1	1	3	3
10/100	Same	Mean	13.28	13.28	13.28	13.28
10/100	Same	Max	100	100	100	100
10/100	Different	Min	0	0	0	0
10/100	Different	Mean	6.74	6.74	6.74	6.74
10/100	Different	Max	20	30	28	28
20/400	Same	Min	0	3	0	3
20/400	Same	Mean	26.645	26.645	26.645	26.645
20/400	Same	Max	400	400	400	400
20/400	Different	Min	0	0	0	0
20/400	Different	Mean	13.36	13.36	13.36	13.36
20/400	Different	Max	158	160	108	108

consecutive rendezvous. For the same number of average potential rendezvous per cycle, a higher variance means that channel matchings occur in bursts, leaving longer gaps between bursts, while the lower variance represents the case when channel matchings are more regularly distributed in time. The lower variance case is better since it would assure that two CRs going online would not be stuck into the long no-rendezvous gaps before a successful rendezvous.

Monte Carlo simulations were made to test the performance of the RAC^{2}E-gQS protocol, for 5, 10 and 20 channels. A total of 10,000 trials (rendezvous cycles) were made for each case for statistical correctness. The simulations were performed for a mean rendezvous cycle duration $T_c = 1$ s and duration of $\tau_{min} = 1 \, \mu$s. This τ_{min} duration roughly maps to a case when we have 10 Msps sampling rate, 1 byte of beacon and beacon reply message lengths and 4-QAM modulation. Different randomization intervals were evaluated, for $k(k = T_c/\Delta T)$ ranging from 1/4 up to 2 with step size of 1/4.

In order to justify the need of randomization introduced by RAC^{2}E, the grid-quorum channel mapping schemes were tested for a scenario of slot synchronized CRs aiming for rendezvous. Slot shifts are likely to occur since both CRs do not start the rendezvous phases simultaneously. Table 4.8 presents the performances of the grid-quorum schemes in terms of the minimum, the maximum and the average number of potential rendezvous per cycle with respect to the slot shifts. As evident slot shifts can cause high time-to-rendezvous (low avg. number of potential rendezvous per cycle) even in the case when both CRs have the same

Table 4.9 Average number of potential RDVs per cycle for the RAC^{2}E-gQS; No.c/s stands for number of channels/slots; Ch. Rank. refers to channel ranking lists

No.c/s	Ch. Rank.	PoP^{RC}	GD^{RC}	PoP^{DC}	GD^{DC}
5/25	Same	13.042	13.042	13.037	13.043
5/25	Different	7.1065	7.1207	7.0994	7.1045
10/100	Same	26.563	26.557	26.554	26.558
10/100	Different	13.409	13.408	13.434	13.356
20/400	Same	53.263	53.243	53.283	53.325
20/400	Different	26.543	26.424	26.515	26.504

channel ranking lists. The different ranking lists and several slot shifts between can result even in no rendezvous between the CRs.

Table 4.9 presents the average number of potential rendezvous per cycle for the RAC^{2}E-gQS protocol, for the *same channel ranking lists* and *different channel ranking lists* of the CRs. It is evident that the case of same channel ranking lists of the both CRs, results in higher average number of potential rendezvous per cycle than the case with different channel ranking lists. RAC^{2}E improves the rendezvous performances of the grid-quorum channel mapping schemes, as evident comparing Tables 4.8 and 4.9 results. The channel matching percentage, calculated as average number of potential rendezvous per cycle divided by the number of slots, is about 52, 26 and 13.25 % for 5, 10 and 20 number of channels, respectively, for the same channel ranking lists case and two times lower for the case with different channel ranking lists.

All inspected grid-quorum channel mapping methods (PoP-RC, GD-RC, PoP-DC, GD-DC), for the particular channel ranking cases and the particular numbers of available channels, provide the same average number of potential rendezvous per cycle. Although most of the methods experience same (or similar) average number of potential rendezvous per cycle, they differ in the inter rendezvous time variance as demonstrated on Fig. 4.17. It presents the dependence of the variance between consecutive rendezvous with the factor of randomization k, for the cases with same and different channel ranking lists and for 5, 10 and 20 channels. Among the grid quorum strategies, the PoP-DC and GD-DC achieve the lowest variance between rendezvous, for the cases with large number of channels, different channel ranking lists and small number of channels, same ranking lists.

Regarding the randomization factor k, it is evident that there is an optimal setting providing the lowest variance between potential rendezvous. The optimal k depends on the number of available channels, the difference between the channel ranking lists and the employed channel mapping method (Fig. 4.17).

The subsection presented a specific rendezvous protocol for control channel management in dynamic ad-hoc environments, mapped into the C-MAC cycle. The RAC^{2}E-gQS protocol handles and facilitates the asynchronous operation to provide better rendezvous performances. Furthermore, it uses specific grid-quorum systems to handle the channel heterogeneity and provide prioritization for better candidate channels in the rendezvous process. The performance analyses [56] prove
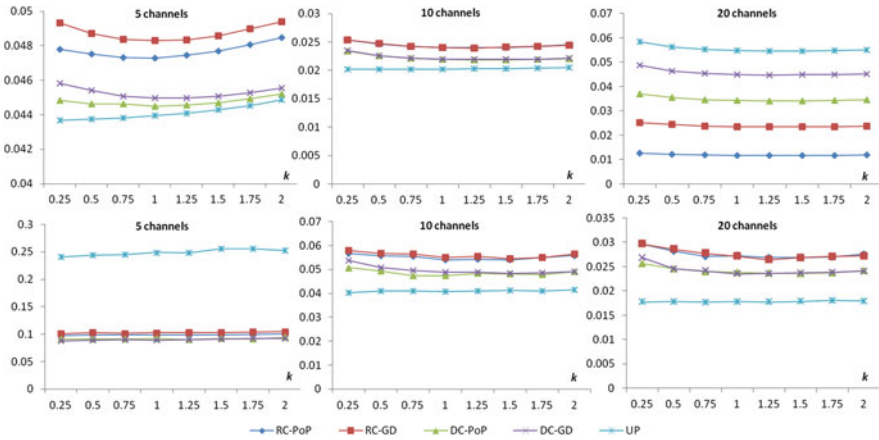
Fig. 4.17 Inter-rendezvous time variance vs. randomization coefficient k, *first row*: same channel ranking case, *second row*: different channel ranking case

the viability of the used grid-quorum schemes for the rendezvous purposes, and the dependence on the randomization introduced by the RAC^{2}E protocol. The results can serve as guidelines for the selection of the most optimal mode of operation of the RAC^{2}E-gQS protocol.

4.6   Concluding Remarks

The Medium Access Control protocols for Cognitive Radio Networks (i.e. the C-MAC protocols) are vital in the process of achieving large spectrum efficiency gain. Utilizing various optimization strategies, the C-MAC protocols strive to maintain the required QoS parameters for the secondary system while providing maximal protection to the primary system. This chapter presented the C-MAC cycle as a generic C-MAC protocol classification and systematization layout, by identifying the spectrum sensing, spectrum sharing and the control channel management as the main generic functionalities. Additionally, the chapter presents brief survey on the state-of-art advances in C-MAC protocol engineering by reviewing existing technical solutions and proposals, identifying their basic characteristics and placing them into the C-MAC cycle layout, with emphasis the C-MAC cycle modularity. It provides overview of large number of technical details concerning the three generic functionalities, (i.e. the radio environmental data acquisition, the spectrum sharing and the control channel management) as the main building blocks of the C-MAC cycle, and related functionality-specific and common aspects. To illustrate the generality of the C-MAC cycle layout, the authors present three C-MAC cycle use cases by mapping a particular C-MAC solution onto a specific generic

functionality. These use cases serve as an illustration of the potential benefits of applying the C-MAC cycle on certain C-MAC protocols in terms of understanding the range of applicability of the protocols and their operational limitations.

References

1. Mitola, J.: Cognitive radio for flexible mobile multimedia communication. Paper Presented at 1999 IEEE International Workshop on Mobile Multimedia Communications, San Diego, Nov 1999
2. Standard for Recommended Practice for Installation and Deployment of IEEE 802.22 Systems, Official IEEE Standard, 28 Sept 2012
3. IEEE 802.11af Standard for Wireless LAN in TV White Space. http://grouper.ieee.org/groups/802/11/Reports/tgaf_update.htm. Accessed 20 Dec 2013
4. EC FP7-248303 project QUASAR. http://www.quasarspectrum.eu/. Accessed 20 Dec 2013
5. EC FP7-216856 project ARAGORN. http://www.ict-aragorn.eu/. Accessed 20 Dec 2013
6. EC FP7-248351 project FARAMIR. http://www.ict-faramir.eu/. Accessed 20 Dec 2013
7. EC FP7-257626 project ACROPOLIS. http://www.ict-acropolis.eu/. Accessed 20 Dec 2013
8. Gavrilovska, L., et al.: Medium access control protocols in cognitive radio newtorks: overview and general classification. IEEE Commun. Surv. Tutor. Accepted 2014
9. Walke, B.H., et al.: IEEE 802 Wireless Systems: Protocols, Multi-Hop Mesh/Relaying, Performance and Spectrum Coexistence. Chichester, West Sussex, England (2006)
10. Krishna, T.V., Das, A.: A survey on MAC protocols in OSA networks. Comput. Netw. J. (Elsevier) **53**(9), 1377–1394 (2009)
11. Domenico, A.D., et al.: (2012) A Survey on MAC Strategies for Cognitive Radio Networks. IEEE Commun. Surv. Tutor. **14**(1), 21–44
12. Ren, P., et al.: A survey on dynamic spectrum access protocols for distributed cognitive wireless networks. EURASIP J. Wirel. Commun. Netw. **60** (2012). doi:10.1186/1687-1499-2012-60
13. Cormio, C., Chowdhury, K.R.: A survey on MAC protocols for cognitive radio networks. Ad Hoc Netw. J. (Elsevier) **7**(7), 1315–1329 (2009)
14. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio application. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009)
15. Denkovski, D., et al.: HOS based goodness-of-fit testing signal detection. IEEE Commun. Lett. **16**(3), 310–313 (2012)
16. Akyildiz, I.F., et al.: Cooperative spectrum sensing in cognitive radio networks: a survey. Elsevier Phys. Commun. J. **4**(1), 40–62 (2011)
17. Aziz, A.M., et al.: A new soft-fusion approach for multiple-receiver wireless communication systems. ETRI J. **33**(3), 310–319 (2011)
18. Ciuonzo, D. et al.: Decision fusion in MIMO wireless sensor networks with channel state information. Paper Presented at 2012 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), Hoboken, June 2012
19. Cordeiro, C., et al.: Cognitive PHY and MAC layers for dynamic spectrum access and sharing of TV bands. Paper Presented at 2006 First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS), Boston, Aug 2006
20. Khan, Z., et al.: Autonomous sensing order selection strategies exploiting channel access information. IEEE Trans. Mob. Comput. **12**(2), 274–288 (2013)
21. Peh, E., et al.: Optimization of cooperative sensing in cognitive radio networks: a sensing-throughput tradeoff view. IEEE Trans. Veh. Technol. **58**(9), 5294–5299 (2009)
22. Mehanna, O., et al.: Blind cognitive MAC protocols. Paper Presented at 2009 IEEE International Conference on Communications, Dresden, June 2009

23. Li, X., et al.: Optimal cognitive access of Markovian channels under tight collision constraints. IEEE J. Sel. Areas Commun. **29**(4), 746–756 (2011)
24. Zhang, Z., et al.: Channel exploration and exploitation with imperfect spectrum sensing in cognitive radio networks. IEEE J. Sel. Areas Commun. **31**(3), 429–441 (2013)
25. Pollin, S., et al.: A distributed multichannel MAC protocol for multihop cognitive radio networks. IEEE Trans. Veh. Technol. **59**(1), 446–459 (2010)
26. Kim, H., Shin, K.G.: Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks. IEEE Trans. Mob. Comput. **7**(5), 533–545 (2008)
27. Su, H., Zhang, X.: Channel exploration and exploitation with imperfect spectrum sensing in cognitive radio networks. IEEE J. Sel. Areas Commun. **26**(1), 118–129 (2008)
28. Hu, D., Mao, S.: A sensing error aware MAC protocol for cognitive radio networks. ICST Trans. Mob. Commun. Appl. **12**(2), e1 (2012). http://eudl.eu/doi/10.4108/mca.2012.07-09.e1
29. Bkassiny, M., et al.: Optimal and low-complexity algorithms for dynamic spectrum access in centralized cognitive radio networks with fading channels. Paper Presented at 2011 IEEE Vehicular Technology Conference (VTC-Spring) 2011, Budapest, May 2011
30. Zhao, Q., et al.: On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance. IEEE Trans. Wirel. Commun. **7**(12), 5431–5440 (2008)
31. Jia, J., et al.: HC-MAC: A Hardware-constrained cognitive MAC for efficient spectrum management. IEEE J. Sel. Areas Commun. **26**(1), 106–117 (2008)
32. Xia, W., et al.: Optimization of cooperative spectrum sensing in ad-hoc cognitive radio networks. Paper Presented at 2010 IEEE International Conference on Communications, Miami, Dec 2010
33. Karami, E., et al.: Cluster size optimization in cooperative spectrum sensing. Paper Presented at 2011 Ninth Annual Communication Networks and Services Research (CNSR) Conference, Ottawa, May 2011
34. Zhao, Y., et al.: Resource allocation in multiuser OFDM system based on ant colony optimization. Paper Presented at 2010 IEEE Wireless Communications and Networking Conference, Sidney, Apr 2010
35. Jiang, X., et al.: Cross-layer design of partial spectrum sharing for two licensed networks using cognitive radios. Wirel. Commun. Mob. Comput. (2013). doi:10.1002/wcm.2345
36. Li, F., et al.: Improved quantum genetic algorithm for competitive spectrum sharing in cognitive radios. Paper Presented at 2011 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, Nov 2011
37. Wang, Z., Zhang, W.: Spectrum sharing with limited channel feedback. IEEE Trans. Wirel. Commun. **12**(5), 2524–2532 (2013)
38. Tang, M., et al.: Nonconvex dynamic spectrum allocation for cognitive radio networks via particle swarm optimization and simulated annealing. J. Comput. Netw.: Int. J. Comput. Telecommun. Netw. **55**(11), 2690–2699 (2012)
39. Tan, L.T., Le, L.B.: Distributed MAC protocol for cognitive radio networks: design, analysis, and optimization. IEEE Trans. Veh. Technol. **60**(8), 3990–4003 (2010)
40. Mahmoud, H., et al.: OFDM for cognitive radio: merits and challenges. IEEE Wirel. Commun. **16**(2), 6–15 (2009)
41. Wilcox, D., et al.: On spatial domain cognitive radio using single-radio parasitic antenna arrays. IEEE J. Sel. Areas Commun. **31**(3), 571–580(2008)
42. Chavez-Santiago, R., et al.: Cognitive radio for medical body area networks using ultra wideband. IEEE Wirel. Commun. **19**(4), 74–81 (2012)
43. Kamruzzaman, S.M.: Dynamic TDMA slot reservation protocol for cognitive radio ad hoc networks. Paper Presented at 2010–13th International Conference on Computer and Information Technology (ICCIT), Dhaka, Dec 2010
44. Cheng, Z.: Cognitive radio based multiple access algorithm for differential frequency hopping network. Paper Presented at 2009 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom), Beijing, Sept 2009
45. Yongwei, H., et al.: Robust multicast beamforming for spectrum sharing-based cognitive radios. IEEE Trans. Signal Process. **60**(1), 527–533 (2012)

46. Tae, W.B.: Capacity and energy efficiency of multi-user spectrum sharing systems with opportunistic scheduling. IEEE Trans. Wirel. Commun. **8**(6), 2836–2841 (2009)
47. Khan, Z., et al.: Modeling the dynamics of coalition formation games for cooperative spectrum sharing in an interference channel. IEEE Trans. Comput. Intell. AI Games **3**(1), 17–30 (2011)
48. Christian, I., et al.: Spectrum mobility in cognitive radio networks. IEEE Commun. Mag. **50**(6), 114–121 (2012)
49. Feng, C., et al.: Cognitive learning-based spectrum handoff for cognitive radio network. Int. J. Comput. Commun. Eng. **1**(4), 350–353 (2012)
50. Zhang, Z., et al.: Self-organization paradigms and optimization approaches for cognitive radio technologies: a survey. IEEE Wirel. Commun. **20**(2), 36–42 (2013)
51. Pawełczak, P., et al.: Performance analysis of multichannel medium access control algorithms for opportunistic spectrum access. IEEE Trans. Veh. Technol. **58**(6), 3014–3031 (2009)
52. Zhao, J., et al.: Distributed coordination in dynamic spectrum allocation networks. Paper Presented at 2005 First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySpan), Baltimore, Nov 2005
53. Liu, S., et al.: Cluster-based control channel allocation in opportunistic cognitive radio networks. IEEE Trans. Mob. Comput. **11**(10), 1436–1449 (2012)
54. Doerr, C., et al.: Dynamic control channel assignment in cognitive radio networks using swarm intelligence. Paper Presented at 2008 IEEE Global Communications Conference (GLOBECOM), New Orleans (2008)
55. Bian, K., et al.: Control channel establishment in cognitive radio networks using channel hopping. IEEE J. Sel. Areas Commun. **29**(4), 689–703 (2011)
56. Romaszko, S., et al.: Asynchronous rendezvous protocol for cognitive radio ad hoc networks. Lect. Notes Inst. Comput. Sci. Soc. Inf. Telecommun. Eng. **111**, 135–148 (2013)
57. Masri, A.M., et al.: Common control channel allocation in cognitive radio networks through UWB communication. J. Commun. Netw. (JCN) **14**(6), 710–718 (2013)
58. Federal Communications Commission (FCC). http://www.fcc.gov. Accessed 20 Dec 2013
59. EC FP7-248303 project QUASAR (2013) Deliverable 4.2. http://quasarspectrum.eu/images/stories/Documents/deliverables/QUASAR_D4.2.pdf. Accessed 20 Dec 2013
60. 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)
61. Torrieri, D.: The radiometer and its practical implementation. Paper Presented at 2010 IEEE Military Communications Conference (MILCOM), San Jose (2010)
62. Shen, B., et al.: Energy detection based spectrum sensing for cognitive radios in noise of uncertain power. Paper Presented at 2008 International Symposium on Communications and Information Technologies (ISCIT), Lao, Oct 2008
63. Mariani, A., et al.: Effects of noise power estimation on energy detection for cognitive radio applications. IEEE Trans. Commun. **59**(12), 3410–3420
64. Rakovic, V., et al.: Cooperative spectrum sensing based on noise power estimation. In: International Symposium on Wireless Personal Multimedia Communications, Atlantic City, June 2013
65. Rakovic, V., et al.: Clustered network coordinated beamforming for cooperative spectrum sharing of multiple secondary systems. Paper Presented at 2011 International Conference on Cognitive Radio and Advanced Spectrum Management (CogART), Barcelona, Oct 2011
66. Chae, C.B., et al.: Network coordinated beamforming for cell-boundary users: linear and nonlinear approaches. IEEE J. Sel. Top. Signal Process. **3**(6), 1094–1105 (2009)
67. Chae, C.B., et al.: Coordinated beamforming with limited feedback in the MIMO broadcast channel. IEEE J. Sel. Areas Commun. **26**(8), 1505–1515 (2008)
68. LTE; E-UTRA; UE radio transmission and reception. 3GPP technical specification 36.101
69. Erceg, V., et al.: TGn channel models. IEEE 802.11 document 802.11-03/940r4 (2004)
70. Pavlovska, V., et al.: Novel rendezvous protocol for asynchronous cognitive radios in cooperative environments. Paper Presented at 2010 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, pp. 26–30, Sept 2010

71. Jiang, J.R., et al.: Quorum-based asynchronous power-saving protocols for IEEE 802.11 ad hoc networks. J. Mob. Netw. Appl. **10**(1/2), 169–181 (2005)
72. Maekawa, M.: A p N algorithm for mutual exclusion in decentralized systems. ACM Trans. Comput. Syst. **3**(2), 145–159 (1985)
73. Romaszko, S., Mähönen, P.: Quorum-based channel allocation with asymmetric channel view in cognitive radio networks. Paper Presented at 2011 6th ACM PM2HW2N Workshop (co-located with MSWiM'11), Miami (2011)

Chapter 5
Dynamic Channel Selection for Cognitive Femtocells

Gustavo Wagner Oliveira da Costa, Andrea Fabio Cattoni,
Preben E. Mogensen, and Luiz A. da Silva

Abstract The ever-growing demand for mobile broadband is pushing towards the utilization of small cells, including metrocells, picocells and femtocells. In particular, the deployment of femtocells introduces significant challenges. First, the massive number of expected femtocells cannot be deployed using the traditional planning and optimization techniques. This leads to uncoordinated deployment by the end-user. Second, the high density of femtocells, including vertical reuse, leads to very different inter-cell interference patterns than the ones traditionally considered in cellular networks. And last, but not least, the possibility of having closed-subscriber-groups aggravates the inter-cell interference problems. In order to tackle these issues we consider the implementation of some aspects of cognitive radio technology into femtocells, leading to the concept of cognitive femtocells. This chapter focuses on state-of-art techniques to manage the radio resources in order to cope with inter-cell interference in cognitive femtocells. Different techniques are presented as examples of gradually increasing sophistication of the cognitive femtocells, allowing for dynamic channel allocation, dynamic reuse and negotiated reuse based on information exchanged with neighbor cells.

G.W.O. da Costa (✉) • A.F. Cattoni
Radio Access Technology Section in Department of Electronic Systems, Aalborg University,
Niels Jernes Vej 12 A6, DK-9220, Aalborg, Denmark
e-mail: gc@es.aau.dk; afc@es.aau.dk

P.E. Mogensen
Department of Electronic Systems, Aalborg University, Aalborg, Denmark

L.A. da Silva
Virginia Tech Research Center – Arlington, 900 N Glebe Road, Arlington, VA 22203, USA
Trinity College Dublin, Dublin 2, Ireland
e-mail: ldasilva@vt.edu

5.1 Introduction

The wireless communication industry faces significant challenges ahead. Projections are that wireless broadband traffic demand may grow as much as 1,000 times in one decade. Delivering such huge capacity growth in such a short period would be already a significant challenge without considering the following aspects which will be faced in this decade:

- It will become even harder to have new spectrum assignments in the spectrum range where transceiver technology is well developed (below 6 GHz). The price of exclusive spectrum licenses may grow considerably.
- The channel coding efficiency achieved e.g. by turbo coding is approaching the theoretical boundary.
- Significantly increasing the number of antennas in a mobile terminal is a major challenge in the frequencies typically used today (below 6 GHz).

Considering all these hurdles for other possible technological developments, two key technologies are gaining increasing attention as candidates to help materializing the view of 1,000-fold increased capacity:

- The massive deployment of small cells, including picocells and femtocells.
- The utilization of dynamic spectrum sharing, empowered by Cognitive Radio (CR) technology.

This chapter combines both concepts, advancing the emerging view of *Cognitive Femtocells (CFs)*, built upon cognitive radio technology to address the interference problems in femtocell scenarios. The presented CF view is: self-optimizing femtocells empowered by limited-complexity Cognitive Access Points (CAPs) capable of serving both legacy devices and cognitive ones.

Typical CR applications are node-centric, i.e., each node may make independent adaptations. In CFs, however, some cognitive decisions are more naturally made by the CAP, also on behalf of the User Equipments (UEs). Using measurement feedback from the UEs, the CAP can analyze the interference scenario for the whole cell. Therefore, UEs may assist on the spectrum sensing, while the CAP can concentrate most of the cognitive functions, coordinating medium and backhaul access.

Femtocells present a perfect use case for CR technology because the fierce interference is disruptive if left unmanaged and the very high number of nodes involved is a major showstopper for traditional centralized solutions. A CF may need to share the spectrum with macrocells and other CF. These issues are summarized in Sect. 5.2. Then, Sect. 5.3 introduces our system model and assumptions used throughout the chapter for analysis and simulations. The performance evaluation assumptions are summarized in Sect. 5.4.

The following sections are the core of the chapter, providing detailed analysis in three levels: single-link potential gain (Sect. 5.5), game-theoretic analysis (Sect. 5.6) and graph-theoretical analysis (Sect. 5.7). After that, different solutions

are presented in Sect. 5.8 corresponding to increasing levels of sophistication of CFs: dynamic selection of channels, independent determination of the frequency reuse and selection of channels and reuse through negotiations. The system level performance of these three methods is compared in Sect. 5.9. Concluding remarks and references to additional resources completes the chapter.

5.2   Overview of Femtocells and Challenges

Femtocells are cost-effective, typically user-deployed, low-power base stations mostly providing indoor coverage. The femtocell concept is enticing due to several potential benefits that it offers to operators and end-users:

- Improvement of indoor coverage and performance. When indoor users are covered by macrocell Base Stations (BSs), the UEs can experience poor coverage due to wall penetration from outdoor to indoor. In contrast, with indoor femtocells the UEs can experience good coverage because they are close to the CAP.
- Offload of the macro-cellular network [3]. Since a large part of the total traffic is generated indoors, femtocells can play an important role in capturing traffic which would otherwise be served by outdoor BS. In this way, the load of macrocells is relieved and the users attached to the macro network can experience better service quality.
- Overall cost reductions [4], especially related to backhaul and site acquisition costs. Notice that if femtocells are user-deployed, then site acquisition costs can be completely avoided, and the end-user may bear the costs of backhaul.

Nevertheless, femtocell deployment is not free of shortcomings. On the contrary, there are plenty of challenges.

Probably, the most important game changer in femtocells is the possibility of having Closed Subscriber Group (CSG) femtocells. In CSG femtocells, the CAP can only serve the UEs which are registered in the allowed list. One must remember that the whole cellular network concept is based on letting the UEs be connected to the serving cell with the most favorable geometry and load conditions. When CSG networks are introduced, these two characteristics are affected. For example, a UE may be forced to connect to a particular cell, while the signal received from another cell is much stronger. In case of macro-femto interaction such unfavorable geometry is known to cause coverage holes for the macrocell, in the near vicinity of the CAP [20]. In the case of femto-to-femto interaction the problem is less dramatic than in a macro-femto interaction. Still, the issue is considerable enough to cause poor performance of the victim UE attached to a CAP and interfered by another CAP [11].

In addition to the CSG issue, there are still many other femtocell aspects which deserve attention. Overall, femtocell deployment poses significant challenges when compared to traditional cellular networks:

- *Massive deployment* – on the same geographical area that was covered by just a few cells in the past, the typical deployment today is much denser, and in the future it may be covered by thousands of femtocells. In fact, even though the femtocell density is still very low there are already more femtocells than macrocells [3]. As the number of cells explodes, planning, optimizing, and managing all cells may become too costly, unless a much higher degree of self-configuration and self-optimization is achieved. Ideally, femtocells should be completely plug and play.
- *Uncoordinated deployment* – As with WiFi Access Point(AP) deployment, it can be expected that most femtocell deployments will be performed by the end users, with the exception of some enterprise solutions. Therefore, the location of a CAP will typically be chosen by the end user. Nothing prevents placement of CAP in locations that will generate unfavorable geometry factors.
- *High density* – in the future, the number of femtocells per km^2 is likely to be high. For this reason, each cell can face a large number of interferers, and the interference footprint can be severely different from the one in traditional cellular deployment.

Such characteristics can lead to scenarios where the interference amongst neighbor femtocells is disruptive. For that reason, femtocell deployment demands some form of interference management [11, 20], especially in CSG deployment. In current femtocell deployments there is little concern about intra-tier interference because of the low deployment density. Some femtocells are even locked to a particular frequency. However, in the long term, as the deployment density increases, the intra-tier interference is likely to become the major challenge for further deployment. This is exactly the problem addressed in this book chapter by presenting CFs.

5.3 System Model and Assumptions

Consider two or more CFs which share a particular spectrum band of interest with a total bandwidth of B MHz. The system bandwidth is divided into K orthogonal channels of B/K MHz. The channelization is common for all CFs. The specific question addressed in this chapter is the following: how can each CF autonomously select a subset of the K channels for transmission in order to maximize the throughput of each CF?

There are no other wireless networks than the CFs accessing the shared band during the time of interest. This assumption holds, e.g. if the spectrum band is licensed, or if the inter-system spectrum sharing problem is solved by another mechanism. All the CAPs and UEs are assumed to be able to transmit and receive over the whole band, i.e. over B MHz. In other words, different classes of UEs are not considered.

It is assumed that the duplexing provides full orthogonality between uplink and downlink directions, such as in a Frequency Division Duplexing (FDD) system or

a frame synchronized Time Division Duplexing (TDD) system. This assumption allows us to consider a single link direction independently. Thus, the B MHz only has to be divided among links in one direction. Hereafter, downlink is taken as the main study case.

Since the channels are considered to be orthogonal, it is implicitly assumed that the receivers can mitigate Adjacent Channel Interference (ACI), whereas Co-channel interference (CCI) is treated as noise by the receivers. The Additive White Gaussian Noise (AWGN) model [14] is used to compute the effect of interference on each link. Each channel is considered to be decoded independently and, therefore, the effects of interference are separate for each channel. The capacity provided by a channel is assumed to be a non-decreasing function of Signal to Interference plus Noise Ratio (SINR). However, a minimal SINR is required for decoding useful data. For example, below that SINR the radios are not even able to synchronize. In addition to that, a maximum implemented Modulation and Coding Scheme (MCS) limits the maximum achievable capacity when the SINR is high. For the purpose of evaluation, the following approximation of Long Term Evolution(LTE) system capacity is used in simulations and numerical examples [18]:

$$
C(\gamma) = \begin{cases} 0 & \text{if } \gamma < \gamma_{min} \\ \frac{B}{K} \, B_{eff} \, log_2 \left(1 + \frac{\gamma}{\gamma_{eff}} \right) & \text{if } \gamma_{min} \leq \gamma < \gamma_{max} \\ \frac{B}{K} \, B_{eff} \, log_2 \left(1 + \frac{\gamma_{max}}{\gamma_{eff}} \right) & \text{if } \gamma \geq \gamma_{max} \end{cases} \tag{5.1}
$$

Here, γ represents the SINR in a linear scale and the other parameters are explained next. Equation (5.1) essentially defines a modified Shannon formula, where the gap between the raw channel capacity and the Shannon bound [22] is the factor γ_{eff}. In essence the parameter γ_{eff} measures how well the system is able to turn SINR into raw capacity. If γ_{eff} is equal to one, then the definition is the same as the Shannon bound. The term B_{eff} accounts for overheads such as guard bands, cyclic prefix and control channels, as well as other adjustments which allow us to approximate γ_{eff} as constant throughout the whole SINR range [18]. γ_{min} is the minimum SINR in order to avoid complete blocking of a receiver. γ_{max} is the SINR at which the maximum MCS can be achieved with very low decoding error probability. Naturally, the channel capacity is scaled for the bandwidth, i.e., B/K MHz. The following parameters are used for performance evaluation: $\gamma_{eff} = 2.0$ and $B_{eff} = 0.56$, which are values adjusted for LTE Single Input Single Output(SISO) [18]. γ_{max} is set in order to achieve the maximum LTE spectral efficiency (after discounting overheads), i.e. 75 Mbps over a band of 20 MHz, for a single spatial stream [15]. The threshold $\gamma_{min} = -7$ dB is a typical value for the minimum SINR so that a UE can synchronize to the serving CAP [2].

There is a control channel where UEs can send measurement reports to the CAP. The UEs are able to estimate the SINR and the sum interference in each channel while decoding useful data. If the interference needs to be discriminated according to the source transmitter, the UEs have to stop decoding incoming data and make dedicated coherent measurements of the received power from each interference
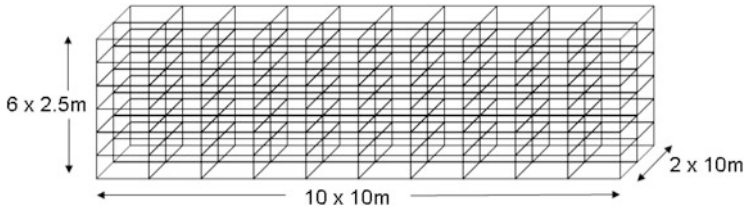
Fig. 5.1 Simulation scenario. A six floor building with 120 apartments

source. These assumptions are in line with LTE receiver capabilities, though not all this information is fed back by default.

The CAP transmits with constant downlink power spectral density over the selected resources. Therefore, the generated interference depends only on which resources are selected for transmission, not on which UE is scheduled.

In order to avoid the joint considerations with different packet schedulers and the inherent difficulties in choosing the feedback information from several UEs, the focus of the chapter is on one UE per CF. For the purposes of practical application, it can be considered that only the information of the UE with the highest backlogged traffic or the UE experiencing the worst SINR is used to make decisions.

Mobility and handovers are not considered. In general, local area mobility is infrequent and relatively slow compared to existing mobile networks. The CFs operate in CSG mode and, consequently, handover considerations within the CF layer are not part of the investigation.

5.4   Simulation Scenario

The scenario adopted for performance evaluation in this chapter is a single building in a stripe format as illustrated in Fig. 5.1. The building has six floors, and the floor height is 2.5 m. In each floor there are 20 apartments and the dimensions of an apartment are $10 \times 10 \times 2.5$ m. Therefore, a total of 120 apartments are simulated. This single stripe scenario is similar to the Third Generation Partnership Project (3GPP) dual stripe scenario [1], except that a single building is considered.

In each simulation drop, some apartments are randomly selected to have a CF deployment. The number of deployed CFs is varied, in order to generate different deployment densities. The number of apartments with a deployed network is set to 1, 10, 20, 30, 60, 90 or 120, corresponding to different deployment ratios: 0.83, 8.3, 16.6, 25, 50, 75 and 100 %. When a CF is deployed, both the CAP and a corresponding UE are deployed within that apartment. Therefore, each apartment may contain a CF or not, and the CFs are randomly distributed over the simulated scenario. The positions of both CAPs and UEs are selected randomly inside the apartment, since a user-deployed scenario is modeled. Each UE always remains

connected to the CAP in the same apartment regardless of the signal and interference conditions, due to the CSG assumption.

In order to achieve statistical reliability, the results were obtained in a Monte-Carlo approach. Each simulation drop corresponds to different random positions for CAPs and UEs. For each deployment ratio and simulated case many simulation drops are performed until samples from 6,000 CFs are available. The results shown throughout the chapter are statistics obtained from 6,000 samples per configuration (simulation case and deployment ratio). This number of samples is enough to have the 95 % confidence interval for the average throughput within ± 2 % of the absolute values, even in the simulated case with largest standard deviation of throughput (reuse 1).

Path loss and shadowing effects are calculated using the propagation model proposed in 3GPP [1], adjusted to a central frequency of 3,450 MHz. This type of propagation model has four components of path loss:

$$\tilde{L} = L_{FS} + L_{LIN} + L_P + L_{FL} \tag{5.2}$$

Here $\tilde{L}$ is the total path loss, L_{FS} is the free space path loss, L_{LIN} is a linear attenuation for a given distance, L_P is the aggregate penetration loss of all crossed walls and L_{FL} is a floor-to-floor loss. All the values are in decibels. The free space path loss can be written as [16]:

$$L_{FS} = 20log_{10}(d) + 46.4 + 20log_{10}(f/5.0) \tag{5.3}$$

Where d is the distance in meters and f is the frequency in GHz. The linear attenuation term models the existence of obstacles, such as furniture, and internal walls within an apartment:

$$L_{LIN} = \alpha d \tag{5.4}$$

The constant α was set to 0.7 dB/m. L_P models the penetration loss in the walls between different apartments and it corresponds to:

$$L_P = n_W L_W \tag{5.5}$$

Where n_W is the number of crossed walls and $L_W = 5$ dB is the penetration loss of a single wall. Finally, L_{FL} models the extra path loss incurred from floor to floor [16]:

$$L_{FL} = 17 + 4(n_F - 1) \tag{5.6}$$

Where n_F is the number of floors between the transmitter and receiver. If the CAP and UE are on the same floor, then L_{FL} is set to zero.

Multipath fading is not explicitly modeled. Shadow fading is modeled by a log-normal distribution (in a linear scale). The shadow fading of each link is calculated

independently, i.e. no spatial correlation of shadow fading is considered. When the CAP and UE are located in the same apartment the log-normal distribution is set to 3 dB standard deviation. Otherwise the standard deviation is 6 dB. A minimum coupling loss of 45 dB was assumed, and the final path loss is set to be at least as large as the free space path loss. Therefore, the path loss incurred in a link is:

$$L = min(\tilde{L} + L_{SF}, L_{FS}, 45) \tag{5.7}$$

Where L_{SF} is the shadow fading of a link, in dB, a value randomly drawn from a normal distribution with the aforementioned standard deviation.

The transmit power of CAPs and UEs was set to 24 dBm. This transmit power is only used over the channels selected by the channel allocation algorithms. The SINR is calculated for each channel using the AWGN model. Look-up tables map the SINR to corresponding throughput values according to a modified Shannon's formula from [18], expressed in Eq. (5.1).

The simulations assume perfectly elastic traffic, also known as full buffer traffic or queue saturation. This type of traffic has the characteristic of perfectly adapting to the physical layer capacity. On the one hand, full buffer traffic provides a worst case situation in terms of generating high interference constantly. On the other hand, assuming perfect elasticity can be a rather idealistic modeling of intermediate layers, such as Transmission Control Protocol(TCP). Nevertheless, the simplistic full buffer traffic assumption allows us to study the problem in isolation of the upper layers' behavior.

The total system bandwidth is 60 MHz, from 3,420 to 3,480 MHz, configured for TDD. The downlink to uplink ratio is set to 7:3, and all CFs are frame aligned. The band is divided into 12 orthogonal channels. In principle, the results should be independent of the method to achieve channel orthogonality, as long as full orthogonality among channels can be attained. The process of coding and decoding is assumed to be independent for each channel. This is true, for example, in LTE Advanced (LTE-A) Component Carriers (CCs). Even though 12 CCs is not a configuration currently in consideration in LTE-A, 12 channels allow us to define precisely reuses 1, 2, 3, 4 and 6 while maximizing co-channel distance. All throughput results are from the downlink, and they were normalized by the maximum theoretical capacity of the system. Hence, a normalized throughput of 100 % means transmission over the whole bandwidth at the maximum system spectral efficiency considering the MCS limitation.

In each snapshot 200 radio frames were simulated, where a radio frame lasts 10 ms as in an LTE-A system. The UE measurements are reported back to the CAP every radio frame. Unless otherwise stated, the throughput results were collected only in the last 50 frames, in order to allow for the convergence of the iterative methods.

5.5 Potential Link Capacity Gain

The first analysis done here is to determine what the potential capacity gain is from a single link point of view. A basic comparison that can be made for a particular link is to evaluate what the capacity is if the transmission is free of interference and what the capacity is when interference is present. In particular, this comparison can be done when the Signal to Noise Ratio (SNR) is very high, such that $SNR \geq \gamma_{max}$. In this case, the received power is enough to achieve the highest MCS in case of interference-free transmission. One should note that the condition $SNR \geq \gamma_{max}$ can often be fulfilled in a dense CF deployment, because, usually, the user will be close to the serving CAP. Using Eq. 5.1, under the condition $SNR \geq \gamma_{max}$, one can see that the capacity provided by one channel free of interference is given by:

$$C_{FREE} = \frac{B}{K} \, B_{eff} \, log_2 \left(1 + \frac{\gamma_{max}}{\gamma_{eff}} \right) \tag{5.8}$$

In an interference-limited condition, a link can experience poor SINR even though the SNR is high. So, one important question is how C_{FREE} compares to the capacity of a channel which is impaired by interference. The answer of course depends on the relative strength of the interference. For low Signal to Interference Ratio (SIR), the capacity of an interfered channel is much lower than C_{FREE}, but for a high SIR the capacity of an interfered channel approaches C_{FREE}. Thus the potential gains that can be provided by mitigating interference depend heavily on the power ratio between desired signal and interference signals.

The trade-off between bandwidth utilization and SINR is an essential feature of the investigated problem. In general, when interference is coordinated it is possible to achieve higher SINR, but the price is that each link has to use fewer channels. For example, with the previous link performance assumptions, if a link has $SIR \approx 12.5\,dB$ when using all channels it will need to have at least 1/2 of the channels totally free of interference in order to break even in total capacity. Conversely, if coordinating interference implies having access to only half the total number of channels, links where the original $SIR > 12.5\,dB$ will certainly experience losses from interference coordination, whereas links where $SIR < 12.5\,dB$ will potentially experience gains from interference coordination, assuming the interference can be mitigated. Since the total capacity depends directly on the number of allocated channels, as described in Eq. (5.1), it is of special importance to analyze the SIR at a reuse 1 condition, i.e. when all channels are reused by all links. The lower the reuse 1 SIR, the highest the potential gains from interference coordination.

5.6   Game Theoretic Analysis

In order to make good channel allocation decisions, CF should not only analyze the potential gains based on the links but also take into account the decision of other CFs. The reason is that the decision of neighbor CFs is interwoven due to interference. The field of Game Theory (GT) is particularly useful to formally characterize situations where decisions are coupled, such as in the problem addressed in this chapter: dynamic channel selection. For this reason, this section provides GT analysis of this problem.

A *game* [10] is any situation where the outcome of the choice made by each decision maker is affected by the choices made by other decision makers. Since the channel selection of each CF affects the decision of other CFs through interference, the dynamic spectrum sharing among CFs can be modeled as a game. In GT, a decision maker is called a *player*. In this spectrum sharing game formulation a player corresponds to a CF. Hereafter, the terms CF and player will be used interchangeably.

A *game in strategic form* Γ is a tuple $\Gamma = (\mathscr{I}, (\Sigma_i)_{i \in \mathscr{I}}, (\Pi_i)_{i \in \mathscr{I}})$ where, $\mathscr{I} = \{1, \ldots, |\mathscr{I}|\}$ is the set of players, Σ_i is the pure strategy space of player i, and it is defined for each player in $\mathscr{I}$. A strategy profile $\mathbf{P}$ is a particular selection of strategies for each player $\mathbf{P} = \langle s_1, \ldots, s_{|\mathscr{I}|} \rangle$, where s_i is a strategy of player i. The utility function $\Pi_i : \mathbf{P} \to \mathbb{R}$ is a real valued function determining the preference of each player over the set of all possible strategy profiles.

A particular CF can have several communication links. It is assumed that the nodes within a CF coordinate themselves to access the medium, providing the functions of duplexing and Multiple Access (MA). The following formulation here deals only with how to share the spectrum among the CFs. Let $\mathscr{K} = \{1, 2, \ldots, K\}$ be the pool of dynamically shared channels. Each player has access to all channels in the pool. Furthermore, the channels are orthogonal, i.e., there is no cross-interference between two channels. The strategy space of each player is the same and consists of all possible spectrum usage masks. In the following, $s_i(k)$ is a binary variable such that $s_i(k) = 1$ if the cell transmits in channel k and $s(k) = 0$ if there is no transmission in that channel. Hence, the spectrum usage mask $\mathbf{s_i}$ can be written as the binary vector:

$$\mathbf{s_i} = \left[s_i^{(1)}, \ldots, s_i^{(k)}, \ldots, s_i^{(K)} \right] \tag{5.9}$$

At least one channel must be selected, so that the CF can operate. Consequently the vector with a null channel selection is not considered to be part of the strategy space. The players only interact with each other by means of interference. The received interference power by player i, in channel k is:

$$I_i^{(k)} = \sum_{\substack{j=1 \\ j \neq i}}^{|\mathscr{I}|} s_j^{(k)} I_{ji}^{(k)} = \sum_{\substack{j=1 \\ j \neq i}}^{|\mathscr{I}|} s_j^{(k)} P_j^{(k)} G_{ji}^{(k)} \tag{5.10}$$

Where $I_{ji}^{(k)}$ is the incoming interference from player j to player i in channel k. $P_j^{(k)}$ is the transmit power player j allocated to channel k and $G_{ji}^{(k)}$ is the path gain between j's transmitter and i's receiver. Equation (5.10) also implies that there is no incoming interference from player j in channel k if that player does not transmit at channel k. Similarly, the received signal power of player i is represented by $S_i^{(k)}$:

$$S_i^{(k)} = \begin{cases} P_i^{(k)} G_{ii}^{(k)}, & \text{if } s_i(k) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5.11}$$

Where $P_i^{(k)}$ is the transmit power used by player i, and $G_{ii}^{(k)}$ is the path gain between the transmitter and receiver in CF i. The utility function of each player is simply the sum capacity provided by all selected channels:

$$\Pi_i = \sum_{k=1}^{K} s_i^{(k)} C_i^{(k)} \tag{5.12}$$

Where $C_i^{(k)}$ is the channel capacity of channel k, and it represents the link level performance of the system. Using the AWGN assumption for the interference effect and Eqs. (5.10) and (5.11), one can write:

$$\Pi_i = \sum_{k=1}^{K} s_i^{(k)} C \left(\frac{S_i^{(k)}}{\sum_{\substack{j=1 \\ j \neq i}}^{|\mathscr{I}|} s_j^{(k)} I_{ji}^{(k)} + N} \right) \tag{5.13}$$

Where $C(SINR)$ is a capacity function such as the one defined in Eq. (5.1), and N is the noise power measured over one channel. Note that $C(0) = 0$ in Eq. (5.1) and thus the case where $S_i^{(k)} = 0$ from Eq. (5.11) does not need special consideration.

The main goal of game theoretic analysis is to predict the behavior of each player when they pursue the maximization of their own utility function. The central concept of non-cooperative GT is the notion of Nash Equilibrium (NE), described next. In GT it is common to denote the vector of strategies of all the players but i as s_{-i}, i.e., $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_{|\mathscr{I}|})$. An NE is a strategy profile where each strategy is the best response to the strategies of the other players. Formally, an NE is a strategy profile (s_i, s_{-i}) where the following condition holds for every i:

$$\Pi_i(s_i, s_{-i}) \geq \Pi_i(\tilde{s}_i, s_{-i}) \text{ for } \forall \tilde{s}_i \tag{5.14}$$

Equation (5.14), explained in words, says that for a fixed strategy from the other players, s_{-i}, if player i selects strategy s_i his payoff will be at least as good as if any other strategy $\tilde{s}_i$ is selected. Since this condition holds for all players, in an NE no player has incentives for making unilateral deviations from the NE strategy profile.

Analyzing Eq. (5.13), a fixed strategy of other players s_{-i} implies a fixed interference level $I_i^{(k)}$ on each channel. For a fixed interference level, the capacity can be maximized using a water filling approach for the distribution of the total power [14]. If the strategy of the other players is to select all channels with equal transmission power, and one assumes flat fading, then water filling application by a player will lead to the decision to select all channels as well. This leads to the conclusion that with such assumptions the strategy profile where all players select all channels is always an NE [8]. While flat fading is not a practical assumption, the average statistics of time-varying channels are usually regarded as the same within a band. This is equivalent to assume an average flat fading behavior over a very long time period.

Such a result is not particularly enticing, since such reuse 1 configuration can be inefficient and have unfair distribution of throughput. Other GT concepts can be applied in order to justify other solutions, such as repeated games [8] and Nash Bargaining Solution (NBS) [23]. A deeper analysis of the two-player version of the spectrum sharing game helps to understand such concepts. In this version of the game, two players i and j can decide their channel allocation among two channels 1 and 2. Thus the strategy space of each player consists of the following 3 vectors: [0 1], [1 0], [1 1]. Recap that the selection [0 0] is ruled out, because an empty allocation would imply that the CF does not operate.

For simplicity, let the transmit power be doubled if both channels are selected. This situation could happen in reality if the transmit power is not limited, but the maximum power spectral density is limited by regulation. With such an assumption the received signal power and received interference power in each channel does not depend on the number of selected channels, but only whether the channel is used or not.

One conclusion that can easily be drawn is that the worst possible outcome is that both players select a single channel and they happen to select exactly the same channel. If the players select a single channel, but they could coordinate their transmissions in order to use different channels, then both players could achieve interference-free transmissions. Iterative Interference Avoidance (IA) methods can effectively rule out the selection of the same channel over time without the need for explicit signaling [7, 17]. Then it is possible to define a very minimal version of the two player spectrum sharing game where each player only has to decide whether to allocate 1 or 2 channels.

Figure 5.2 shows numerical examples for this simplified CF game. The links have very high SNR, a common condition in CFs, but they are interference-limited. The capacity is calculated using Eq. (5.1), with parameters $\gamma_{eff} = 2.0$, $B_{eff} = 0.56$ and bandwidth per channel of 1 MHz. In Fig. 5.2a, the resulting game is the well known prisoners' dilemma [10]. The prisoners' dilemma has only one NE: both players allocate all channels. However, the capacity of each player could be increased if they both used a single channel. This type of game is widely used as one example where the lack of coordination leads to an outcome that is undesirable to both players (because the NE is not Pareto optimal). Starting from the orthogonal allocation,
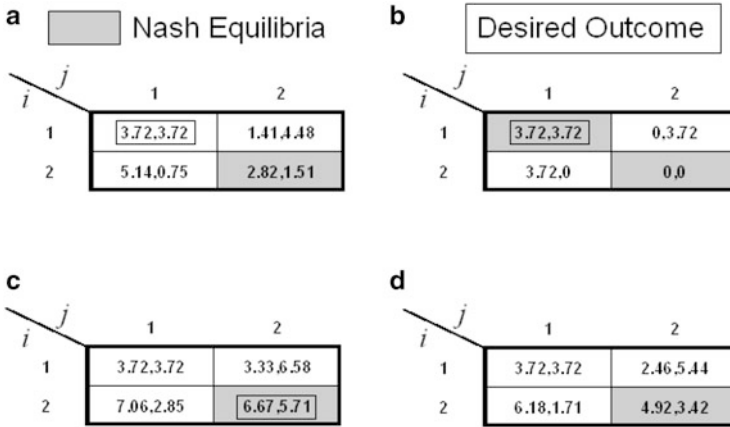
Fig. 5.2 Numerical examples for the game described above. The expected outcome NE is compared with the desired one. (**a**) A prisoner's dilemma SINR = 10 and 5 dB. (**b**) A dummy game SINR = −15 and −10 dB. (**c**) A game where full reuse is highly desirable, SINR = 25 and 20 dB. (**d**) A game where the desired outcome depends on the network optimization criteria. SINR = 17 and 12 dB

either player sees benefit in also allocating the second channel. Notwithstanding, if both players choose to allocate the second channel, they will both be worse off than if they cooperate by selecting a single channel. Furthermore, in such a game the orthogonal allocation is the NBS of the game. On a repeated game, the orthogonal allocation becomes an NE [10], because reuse 1 is Pareto dominated by the orthogonal allocation. Thus, in case the interference interaction between the two CFs is long enough there are incentives to make the orthogonal allocation.

Figure 5.2b shows a case where the interference coupling is so large that the transmitters completely block the receivers in case of overlapping allocation. This is called a dummy game, because the blocked player decision cannot influence its own utility at all. In this game, where both players are blocked, cooperation is expected since the orthogonal allocation is also an NE. In reality, though, such observation could only be made if both transmitters know they are blocking the receivers from each other. Such knowledge may not be available in reality. If only one of the players is blocked, it cannot leave the blocking situation if the other player chooses to transmit in both channels. But the player which is not blocked would have incentives to transmit in both channels. Ultimately, the blocked player depends on the kindness of the other player in order to be able to transmit anything. This is a reason why spectrum etiquettes need to be applied.

When the SINR is sufficiently high, the reuse of both channels is much more efficient than orthogonal allocation. One example is shown in Fig. 5.2c. In this case, the desired outcome matches the NE of the game. Thus, a greedy strategy selection is not always destructive. There are also intermediate cases as illustrated in Fig. 5.2d. In such cases, the definition of a desired outcome is blurry because the orthogonal

allocation provides better outage performance whereas the reuse is more efficient in terms of sum throughput. These examples show the multiple nature of the problem. The players should realize which type of game they are playing before making a decision between orthogonal allocation and reuse of resources. This is not a trivial task to do only with local information (implicit coordination case). In addition to that, for more than 2 players the interaction between each pair of players may be rather different, e.g., corresponding to one of the 4 cases exemplified in Fig. 5.2. The interaction of many CFs is further analyzed in Sect. 5.7 in light of graph theory.

5.7 Graph Theoretic Analysis

In this section, the simulation scenario is analyzed with the aid of graph theory concepts. A *graph* is a mathematical abstraction useful to express a set of objects and the relationship among them. A graph can be either directed or undirected depending on whether the relationship is asymmetric or symmetric. An *undirected graph* $G = (V, E)$ is defined by a set of *vertices*, V, and a set of edges E. Each edge is associated with a pair of vertices, and the presence or lack of an edge in the graph defines the relationship for that pair of vertices. A *directed graph* $G = (V, A)$ is defined similarly; the only difference is that the relation between two vertices is not symmetric. Thus, for each pair of vertices, v_1 and v_2, a directed graph has up to two *arcs*. One arc is directed from v_1 to v_2 and the other arc is directed from v_2 to v_1. By contrast, in an undirected graph an edge between v_1 and v_2 does not make any distinction of direction. A *weighted graph* has weights associated with each edge or arc. Therefore, weighted graphs are suitable to model problems where there is different intensity of relationship among the vertices.

Such graph definitions can be applied to model problems such as the channel assignment problem [19]. A wireless link or a cell corresponds to a vertex. Then, the potential interference among links or cells can be characterized as an edge. As one example, let us consider the scenario shown in Fig. 5.3a, which corresponds to one floor of the simulation scenario. The arrows illustrate the SINR of the victim if the channel is shared only with the interferer at the other end of the arrow. This scenario can be abstracted as the weighted directed graph of Fig. 5.3b. Clearly, from such an abstraction, the interference relationship between each pair of cells can be quite asymmetric. Also, by comparing Fig. 5.3a, b, one can relate very low SINRs to the proximity of the interferer and the distance to the serving CAP. These two characteristics are due to user-deployment and CSG assumptions.

If one wants to express the conflicts, regardless of direction, a possible way is to select the most restrictive SINR, i.e. the lowest, to be an edge on a new graph. This process is illustrated by Fig. 5.3c where the directed graph of Fig. 5.3b is transformed into an undirected graph.

If one wants to mitigate all the interference, one should be able to assign orthogonal frequencies to each of the vertices of the graph which share an edge (highly interfered), while CFs which do not share an edge in such a graph should
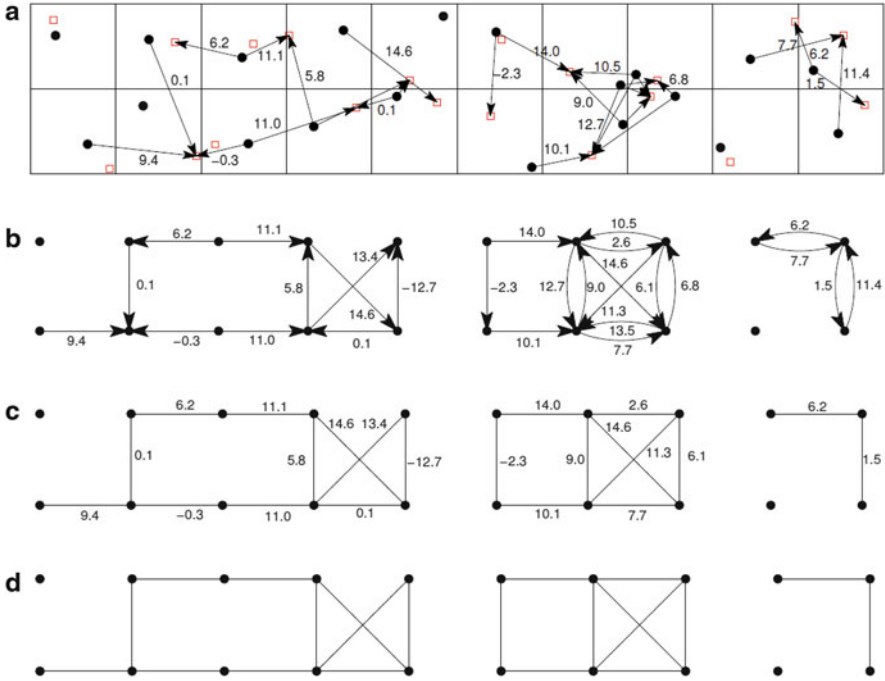
Fig. 5.3 Illustration of how a CF interference scenario can be abstracted to a graph. (**a**) Example of simulated scenario (one floor). The *arrows* point from the interferer (CAP) to the victim (UE). The value near each arrow is the SINR of the victim if the channel is shared only with that interferer. Only SINR values below 15 dB are shown for clarity. (**b**) Graph abstraction of the inter-cell interference from (**a**). Each wireless communication link can be abstracted as a node in a graph. The arcs can represent some conflict, as for example, SINR below 15 dB. (**c**) Graph of (**b**) where the direction was removed, by selecting the arc with the lowest SINR, to become an edge. (**d**) Edges of the graph where SINR is below 15 dB in at least one direction

be able to reuse the spectrum. Naturally, the difficulty is on defining what highly interfered means. This is illustrated in Fig. 5.3d, where the weights of the graph of Fig. 5.3c can be removed by applying an SINR threshold. For different thresholds the density of the graph is very different.

An undirected graph, like the ones shown in Fig. 5.3d, is a simple way of representing whether there is a conflict between each pair of CFs. One relevant question is: how many channels (colors) are needed in order to assign one channel to each CF (vertex) so that no conflict remains? This problem is known as graph coloring [2]. In the case of a channel assignment problem, one channel corresponds to one color. A *proper coloring* is an assignment of a color to each vertex in such a way that no pair of vertices sharing an edge uses the same color. A graph is said to be *k-colorable* if k is the minimum number of colors needed to make a proper coloring.

A **lower bound** on the number of colors (channels) needed for a conflict free assignment can be determined by cliques. A *clique* is a subgraph such that all possible pair of vertices in that subgraph share an edge. Such a definition excludes vertices with no incident edge, i.e., isolated vertices. Large cliques are, by definition, composed of smaller cliques. For example, a clique of size 3 is necessarily composed of 3 cliques of size 2. A *maximal clique* is a clique which is not included in any larger clique. The definition of maximal clique is important because it defines the densest spectrum reuse which can be achieved in a particular area while keeping an orthogonal channel allocation in CFs with a relevant interference coupling. Notice that the definition of a maximal clique size is local. In principle, nothing should preclude, for instance, an isolated vertex (CF) to allocate the whole spectrum.

Note that each vertex of a graph can be part of multiple maximal cliques. For a particular CF, the maximal clique with highest cardinality determines the maximum number of channels that CF can have while: (1) Solving all interference conflicts in the clique by attaining an orthogonal allocation within the clique, and (2) assigning an equal share of channels for each CF in each clique whenever possible. Therefore, the highest cardinality among the maximal cliques is an interesting figure for analysis. This metric is shown in Fig. 5.4 for the investigated scenario, where the threshold to have an edge on the graph was set to 12 dB SINR. This value was chosen because it is roughly the SINR where half the maximum capacity is achieved. As expected, as the deployment ratio increases the distribution shifts towards having more maximal clique sizes of larger cardinality. Nonetheless, even at very high deployment ratio, the vast majority of interference conflicts can be locally solved by dividing the spectrum into 2 or 3 parts. What about the very few cliques of size 4 and 5? Are they always a concern? While it is true that more interference can be avoided by going to such sparser reuses, this is not necessarily a desirable goal.

The trade-off between mitigating all interference or not in larger cliques is easier to understand with an example. In Fig. 5.5, the clique of size 4 from Fig. 5.3b was isolated from the rest of the network. Then, Fig. 5.5 shows the throughput achieved and the residual interference for reuse 1, 2 and 4. In the reuse 1 configuration, all the conflicts translate into summed interference. It can be seen that by mitigating all interference, all CFs can achieve 25 % of the maximum system throughput since they can use 1/4 of the channels at large SNR. This is actually a loss for one of the cells. In reuse 2, most interference is mitigated, but there is still residual interference between some links. The worst case throughput is reduced compared to reuse 4. However, in this scenario, reuse 2 provides a substantial increase of average throughput compared to reuse 4 configuration, because all cells have access to twice the bandwidth. Noticeably, the lower left cell loses even more capacity compared to reuse 1 and reuse 4. Assuming the default game theoretic behavior, and each cell
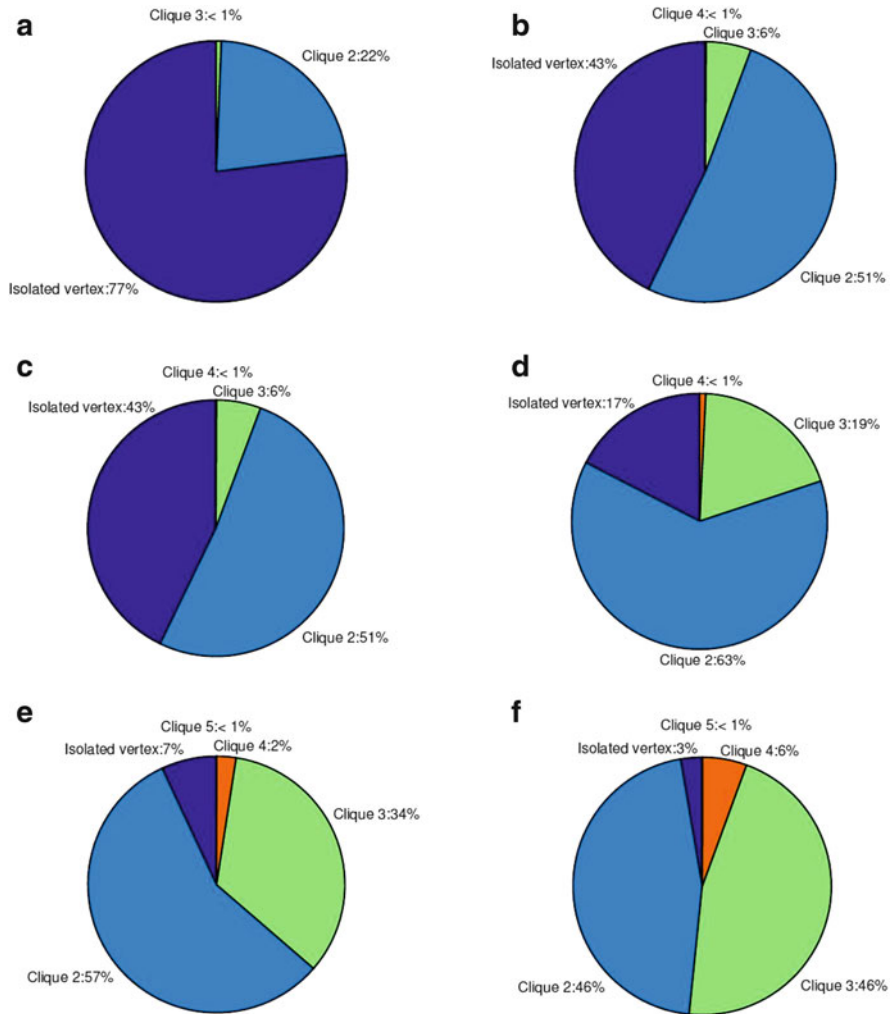
Fig. 5.4 These graphs show the distribution for the following metric: for each CF the maximal cliques are analyzed. Then, the maximal clique with highest cardinality for that CF is taken as a sample. Isolated vertices are also included. Each graph corresponds to a different deployment ratio. (**a**) 8.3 % deployment ratio (10 networks). (**b**) 16.6 % deployment ratio (20 networks). (**c**) 25 % deployment ratio (30 networks). (**d**) 50 % deployment ratio (60 networks). (**e**) 75 % deployment ratio (90 networks). (**f**) 100 % deployment ratio (120 networks)

maximizing its own capacity, such a cell would never accept this reuse 2 solution. From an engineering point of view, however, reuse 2 is an attractive solution to the problem in Fig. 5.5.
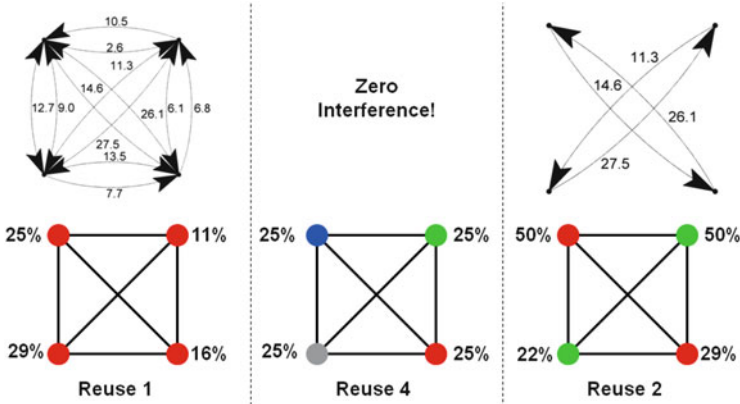
Fig. 5.5 Example of interference scenario with 4 CFs. The interference can be fully mitigated by a completely orthogonal allocation (reuse 4), but a partially orthogonal allocation (reuse 2) may lead to higher average throughput

5.8 Distributed Methods for Channel Allocation

This section gives a brief overview of three different methods for dynamic channel allocation in CFs. These methods have different degrees of complexity and they are examples of practical solutions to the issues discussed in the previous sections.

5.8.1 Dynamic Channel Allocation

The dynamic channel selection method described in [7] attempts to minimize interference for a given reuse configuration. Hereafter, this method is referred as Distributed Inter-cell Interference Coordination (D-ICIC), the title of the original paper. D-ICIC assumes that each cell demands a specific amount of resources. Here, the demanded number of channels m is the same for all cells. The implementation chosen here is a probabilistic version of the best-reply dynamics [9] and slightly simplified version of [7]. Each CF updates its channel allocation autonomously according to a given probability ϵ, which was set to 50 %. If the CF decides to update its allocation, then it selects the least interfered channels. This process is guaranteed to converge if the interference is symmetric between each pair of CFs, because it can be modeled as a potential game [7].

D-ICIC can be used to define a target frequency reuse r, while using a dynamic channel allocation. The reuse is defined as the total number of channels on the band, 12, divided by the amount of demanded channels, m.

D-ICIC is summarized in the following algorithm.

Distributed Inter-cell Interference Coordination (D-ICIC)
Parameter – Target frequency reuse r

  # Some constants are assumed to be given.
  # K is the total number of channels (e.g. 12).
  # ϵ is the status quo review probability.
  $m \leftarrow K/r$ # m is the target number of channels.
  for each radio frame **do**
      $v \leftarrow U(0, 1)$ # Sample a random number between [0,1].
      # With probability ϵ
      if $v < \epsilon$ **then**
          Update interference measurements per channel.
          Order the channels in terms of increasing interference.
          Select the m least interfered channels.
      end if
  end for

5.8.2 Dynamic Reuse Selection

Timeout Based Reuse Selection (TBRS) is a distributed approach, proposed in [6], that tries to discover which is the densest frequency reuse possible in a particular area. Starting from a particular frequency reuse each CF attempts several steps for interference minimization, i.e. IA steps. Such steps may loop, or even in a stable situation the interference may be unbearable with the current number of allocated channels. If either situation is detected by a CF, it will reduce the number of allocated resources.

The number of IA steps which a CF attempts before changing the target reuse is controlled by a timer. Since a fully distributed operation is assumed, each CF has its own timer, and the effective reuse is flexible. For example, this timer can be a deadline to mitigate interference or a countdown timer. The latter approach is taken. Every time the SINR is detected to be below the threshold T in at least one used channel, the countdown timer is decreased; otherwise it remains halted. If the countdown timer expires and interference is still not mitigated a sparser reuse is attempted. Hence, this approach is named *Timeout* Based Reuse Selection (TBRS).

The selection of the frequency reuse can be seen as a state machine composed of R states, where R is the maximum frequency reuse to be supported. Considering only integer reuses, the state machine has only a few states where the reuse of resources is pre-defined for each state: $K/1$, $K/2$, $K/3$ and so on. Notice that the state directly corresponds to the reuse a particular CF is attempting, i.e., in general reuse r is attempted in state r.

The feasibility of a particular reuse is tested by verifying the presence of severe interference or, equivalently, a low SINR. When the interference is *sustained* and the countdown timer expires, then the CF moves onto the next state. Similarly, if low interference is sustained for long enough, then the algorithm may fall back to a denser reuse. This can be implemented by a second timer which ticks when low interference is detected on channels which are not currently in use. In such

implementation the same timer is used to tick in both directions. For example, if the *timeout* is set to 20 frames and the counter reaches 0 the CF will move to a sparser reuse. If the counter reaches 40 the CF will fall back one state to a denser reuse. Note that the IA part of TBRS is essentially the same as in D-ICIC.

The complete method is described as follows.

Timeout Based Reuse Selection (TBRS)
Parameters – SINR threshold T and maximum stage R

  # Some constants are assumed to be given.
  # K is the total number of channels (e.g. 12).
  # *timeout* - number of frames to expire the timer.
  # ϵ is the status quo review probability.
  $n \leftarrow K$ # n is the current amount of channels.
  $r \leftarrow 1$ # r is the current stage.
  $c \leftarrow timeout$ # c is the countdown/countup timer.
  for each radio frame **do**
      Update SINR measurements per channel.
      if any allocated channel has $SINR < T$ **then**
          $c \leftarrow c - 1$
      else if $r > 1$ **then**
          if $K/(r-1) - n$ non allocated channels have $SINR > T$ **then**
             $c \leftarrow c + 1$
          end if
      end if
      if $c = 0$ **then**
          if $r < R$ **then**
             $r \leftarrow r + 1$ # Sustained interference. Use sparser reuse.
             $n \leftarrow K/r$
             $c \leftarrow timeout$
          end if
      else if $c = 2 * timeout$ **then**
          if $r > 1$ **then**
             $r \leftarrow r - 1$ # Interference vanished. Use tighter reuse.
             $n \leftarrow K/r$
             $c \leftarrow timeout$
          end if
      end if
      $v \leftarrow U(0, 1)$ # Sample a random number between [0,1].
      # With probability ϵ
      if $v < \epsilon$ **then**
          Update interference measurements per channel.
          Order the channels in terms of increasing interference.
          Select the n least interfered channels.

      end if
  end for

5.8.3 Negotiated Reuse Selection

The two algorithms described so far, D-ICIC and TBRS, do not consider any type of inter-cell control signaling. This a highly desirable characteristic from a system complexity perspective, which also allows new decisions of channel allocation to be made very quickly and dynamically. This is akin, for example, to the fast and independent decisions we can make while driving, such as changing lanes. In other situations, we may take longer for decisions but we exchange important information, we make requests and we negotiate compromises with other involved parties. Such a process is closer to the algorithm presented in this section.

This framework, named Self-Organizing Coalitions for Conflict Evaluation and Resolution (SOCCER) [12], is composed of these parts:

1. Measurements acquisition and exchange.
2. Conflict evaluation at the critical decision points.
3. The rules which guide the conflict resolution and channel redistribution.

The CAP can establish cooperative sets via multilateral agreements. Once established, the members of a cooperative set will keep orthogonal allocation during the valid period of the agreement. As such, a CAP may be part of no, one or several cooperative sets at the same time.[1] The channel allocation of a particular CF will be orthogonal to all other CFs with which it cooperates. The participation in a cooperative set is assumed to be binding. Once agreed, the participating CAP must respect the agreement, and their packet schedulers shall abide to the imposed restrictions.

In real wireless communication systems, acquiring and exchanging measurements is a major task, so the required information needs to be kept to a minimum. The information assumed in SOCCER is relatively simple: the pairwise characterization of incoming and outgoing interference, known as Background Interference Matrix (BIM). Such information is also the basis for other algorithms proposed for LTE-A [11, 13]. A graphical visualization of BIMs is present in Fig. 5.3a.

The next step is for each CF to evaluate incoming and outcoming BIM values in order to characterize the conflicts among CFs. The intensity of interference coupling (low BIM values), number of conflicts and direction are all important, but the graph theoretical analysis in Sect. 5.7 showed that the intensity is the most important factor. This motivates us to characterize the strength of the interference coupling between a pair of CFs. Such characterization of intensity leads to the definition of *strong bond*. Conceptually, if a strong bond is present between two CAPs implies that mutual cooperation by means of an orthogonal allocation is deemed beneficial. Conversely, in the absence of strong bond the two CFs will regard reusing the same channels as more attractive. Here, two different alternative definitions for

---

[1]Originally, the term *coalition* was used instead of cooperative set. However, the term was avoided here in order to prevent confusion with the usual definition of coalition in GT, which typically only allows each player to be member of a single coalition [21].

strong bond are considered: outage oriented (*Max-min* strong bond) and average throughput oriented (*Max-sum* strong bond).

During bootstrap or when starting a new session, a CAP needs to determine which neighboring CAP should be considered as candidates for cooperation. Each candidate CAP should fulfill two conditions: it has an active session and it shares a strong interference bond with the new entrant. If there are no candidates for cooperation, the solution is trivial: the new entrant CAP can reuse all channels.

If candidates are found the new entrant communicates with them with a simple protocol for the formation of a cooperative set:

1. The new entrant sends a Coordinated Transmissions Request (CTR) message to each candidate, which contains a list of all potential participants in the new cooperative set.
2. Each candidate answers with a Coordinated Transmissions Reply(CTY) which indicates channel allocation restrictions implied by existing cooperative sets. In addition to that, the message indicates whether any of the candidates already form a cooperative set among themselves.
3. The new entrant collects all the information from CTY messages. Then it can calculate the new allocation, which is included in a Coordinated Transmissions Acknowledgment(CTA) message sent back to the candidates. This latter message confirms the formation of the cooperative set which then is completely formed. The new allocation can then take effect.

The rules to allocate the channel can be graphically seen in [12], and they are summarized in the following algorithm, which shows a full pseudocode for SOCCER.

5.9 Results and Discussions

The results in this section were generated according to the simulation scenario described in Sect. 5.4. The comparison is made in steps. First, TBRS is compared to reuse 1 for a large set of parameters. Then, some special cases of TBRS are compared to D-ICIC. Finally, the two variants of SOCCER are compared to the TBRS parameters which provided the best outage and the best average throughput.

TBRS was simulated for several combinations of the two algorithm parameters: R, the number of states (maximum reuse) and the SINR threshold T. The maximum reuse R was varied from 2 to 4 and the SINR threshold T was varied for all the values 0, 5, 10, 15, 20 and 25 dB. The notation TBRS(R,T) is used to summarize the results. Note that reuse 1 can be represented as TBRS(1,T).

Figures 5.6 and 5.7 show TBRS results at 100 % deployment ratio using reuse 1 as a baseline. Reuse 1 fails to provide any performance at all for more than 5 % of the cells. The interference is so strong that not even synchronization is possible. The dependency of outage performance on the maximum stage is quite clear. More stages allow the network to split the resources until interference can mitigated by

---

SOCCER – Self-Organizing Coalitions for Conflict Evaluation and Resolution

   # i is this CF. Other CFs are j or k.

   # n_i, n_j and n_k refer to the number of allocated channels.

   # K is the total number of channels (e.g. 12).

   # BIM_{thr} is the threshold in order to exchange BIMs.

   for each neighbor j **do**

      Measure DL$_{\{i\}\leftarrow\{j\}}$

      if DL$_{\{i\}\leftarrow\{j\}} < BIM_{thr}$ **then**

         Send DL$_{\{i\}\leftarrow\{j\}}$ to j

         Request DL$_{\{j\}\leftarrow\{i\}}$ from j

      end if

   end for

   for each new session **do**

      Select nc candidates.

      if $nc = 0$ **then**

         Allocate all K channels.

      else

         Send CTR to the nc candidates.

         Receive CTY from the nc candidates.

         if $nc = 1$ **then**

            if The candidate j is not part of any cooperative set **then**

               Allocate $K/2$ channels to i, and the other $K/2$ channels to j.

            else

               If $n_j > K/2$ remove channels from j until $n_j = K/2$.

               Allocate to i all channels which are not allocated to j.

            end if

         else if $nc = 2$ **then**

            if j and k already form a cooperative set of size s **then**

               if $s = 2$ **then**

                   Reduce their allocation to $n_j = K/3$ and $n_k = K/3$

                   Allocate to i all channels which are not allocated to j or k.

               else if $s = 3$ **then**

                   Allocate to i all channels which are not allocated to j or k.

               end if

            else

               if Their allocations are compatible with reuse 2. **then**

                   $r \leftarrow 2$

               else

                   $r \leftarrow 3$

               end if

               Remove channels from j and k until $n_j = K/r$ and $n_k = K/r$.

               Allocate to i all channels which are not allocated to j or k.

            end if

         end if

         Send CTA with the updated allocation to the nc candidates.

      end if

   end for
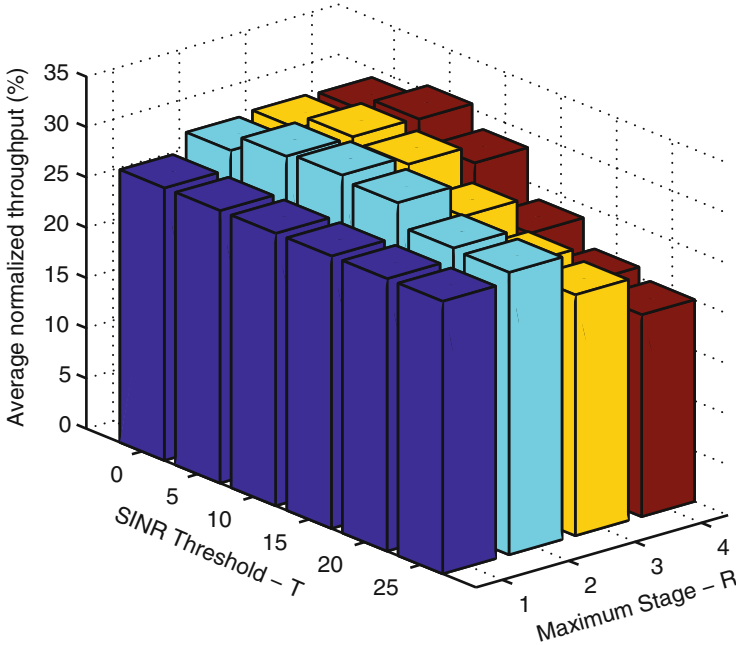
Fig. 5.6 Average throughput performance for 100 % deployment ratio with TBRS algorithm

IA steps. The dependency on the threshold is mostly visible for $R = 3$ and $R = 4$. Instead, if only two stages are allowed, there is not enough room to mitigate interference completely in such a dense scenario. This can be related to Fig. 5.4f, which shows that at 100 % deployment ratio nearly 52 % of the cells are part of cliques of size 3 or 4 (for an edge defined by SINR below 12 dB). A very large SINR threshold reduces the average performance, and some intermediate values, around 5–10 dB allow performance gains in terms of increased average throughput.

In order to understand the benefit of selecting how many channels can be used instead of pre-fixing the demand, one can compare D-ICIC to TBRS with a high threshold. TBRS(R,∞) always ends up in the maximum reuse R, since SINR is finite (there is always noise). Thus, TBRS(R,∞) would generate the same allocation as D-ICIC with $r = R$. If the SINR threshold was set to be exactly the SNR, then one is striving for zero interference. So, as long as the interference is detectable TBRS(R,SNR) would also keep on splitting the band, i.e., moving to the next stage until either reuse R is achieved or all interference can be avoided with a reuse less than R. More interestingly, the comparison of D-ICIC and TBRS(R,25) yields fruitful insights. The average throughput results of such a comparison are summarized in Fig. 5.8.

The main drawback of capping the maximum spectrum utilization becomes evident: low average performance at a low deployment ratio or at low traffic loads. Since D-ICIC assumes a fixed bandwidth utilization, the CFs cannot exploit
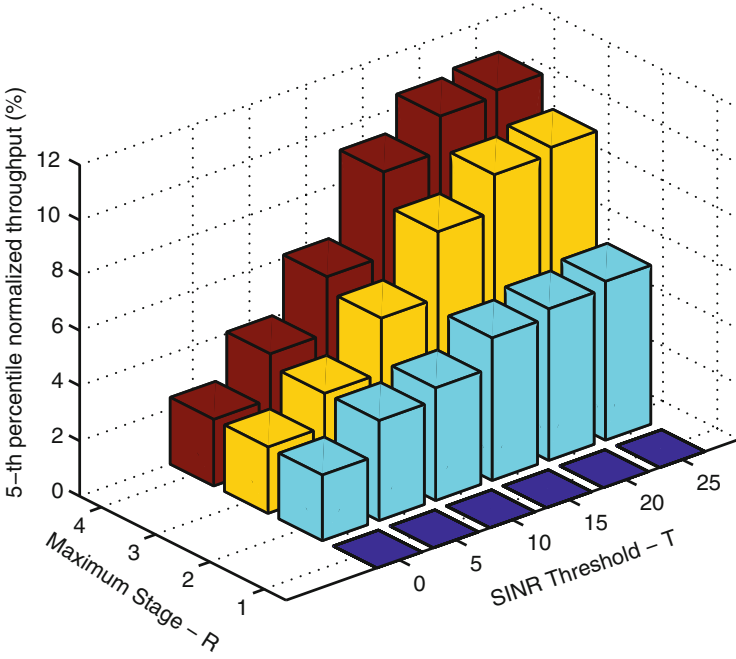
Fig. 5.7 Fifth percentile of throughput performance for 100 % deployment ratio with TBRS algorithm
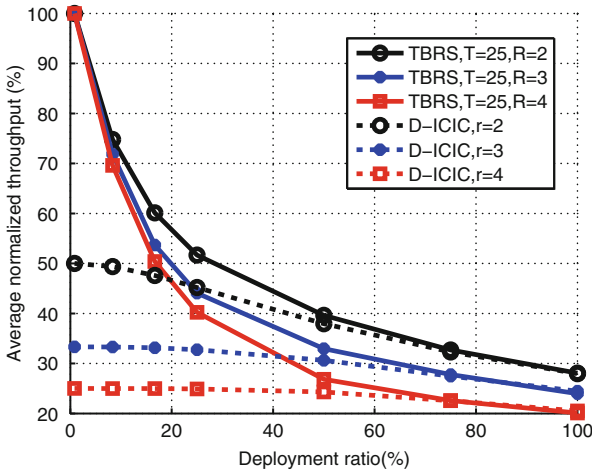
Fig. 5.8 Comparison between TBRS and D-ICIC average cell throughput performance for different deployment ratios

Fig. 5.9 Comparison
between TBRS and D-ICIC
5th percentile throughput
performance for different
deployment ratios
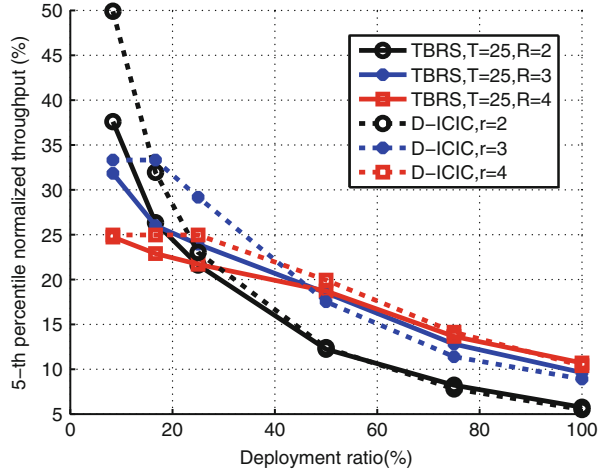
opportunistic reuse even if the interference is virtually non-existent. On the other
hand, TBRS adapts to such situations by keeping denser reuses. As the deployment
ratio increases, TBRS will make the reuse more sparse in order to mitigate
interference.

In nearly all deployment ratios, D-ICIC and TBRS(R,T) perform closely in terms
of the 5th percentile of the distribution, as shown in Fig. 5.9. One exception is
actually when a single network is deployed per building, since TBRS uses the
whole spectrum in this case, but D-ICIC is hard limited. This case was omitted
from Fig. 5.9 so that the differences among the curves could be seen more clearly. It
can be concluded that in this scenario it is better to impose a strong measurement-
based limitation for spectrum access (e.g. 25 dB SINR) rather than hard-limiting the
number of selectable channels. Referring back to Eq. (5.8), at 25 dB the maximum
MCS is reached. Thus, a CF which is already operating at more than 25 dB SINR
in the selected channels will see no benefit in further splitting the spectrum. Notice
that in the studied scenarios SNR $\geq$ 25 dB.

The best average throughput performance of TBRS was achieved with $T = 10$
and $R = 2$, while TBRS(4,25) reached the best outage throughput with 100 %
deployment ratio. These two settings are compared to SOCCER with max-sum
strong bond and SOCCER with max-min strong bond in Figs. 5.10 and 5.11.
SOCCER max-sum and TBRS(2,10) presented very similar performance of average
throughput for all deployment ratios. Actually, the difference between the two
methods in this Key Performance Indicator (KPI) should not be deemed as statis-
tically significant given the confidence interval of the simulations. Nevertheless,
for the same average throughput, SOCCER max-sum can provide better outage
performance at the lower deployment ratios. Thus, SOCCER max-sum gives
preferable outcomes.

The real advantage of exchanging information and negotiating the reuse becomes
clear when comparing SOCCER max-min to the other alternatives. Not only

Fig. 5.10 Average throughput comparison between one explicit coordination method (SOCCER) and another based on implicit coordination (TBRS)
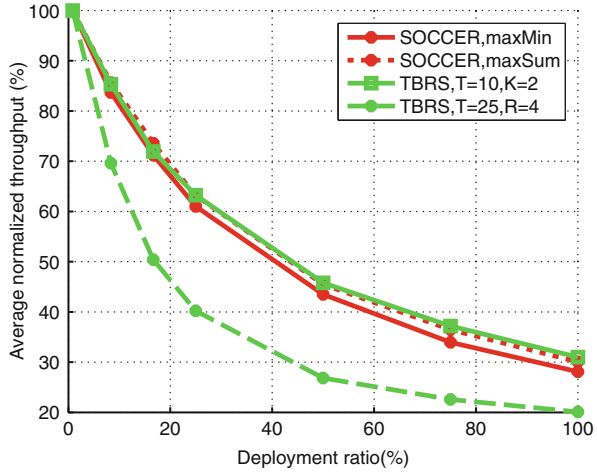
Fig. 5.11 Fifth percentile of throughput comparison between one explicit coordination method (SOCCER) and another based on implicit coordination (TBRS)
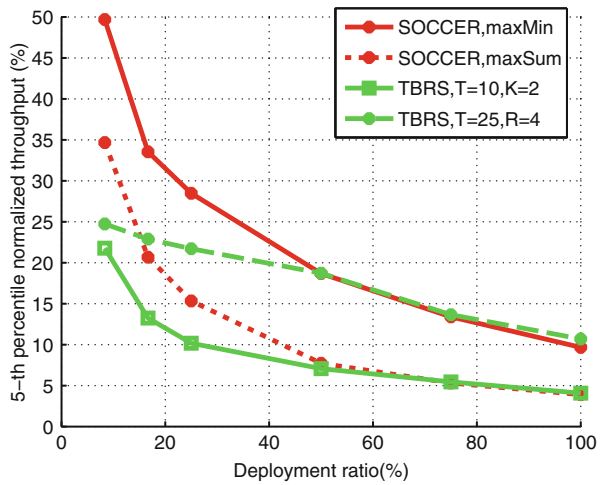
SOCCER max-min provides very high outage performance in all deployment ratios but it achieves such goal with very little average performance sacrifice. Therefore SOCCER max-min shows the best trade-off of all methods discussed in this chapter. In contrast, TBRS(4,25) can only achieve a higher outage performance at a 100 % deployment ratio at the cost of much lower average throughput and without the same degree of adaptivity for all deployment ratios. In fact, the topology knowledge acquired by BIM exchanges let SOCCER max-min adapt to any clique distribution shown in Fig. 5.4, mitigating all cliques of cardinality 2 and 3 and reducing most interference in cliques of cardinality 4.

While this section has shown the benefit of significant superior performance of a method based on explicit coordination, it is important to note that such performance is not the only consideration for practical implementation. Due to signaling delays,

explicit coordination methods cannot adapt as fast as methods based on implicit coordination. Therefore, methods based on implicit coordination may be needed in order to cope with interference fluctuations at small time scales, caused due to mobility and traffic variations. A combination of implicit coordination (fast adaptations) and explicit coordination (slow adaptations) is likely to be the most appropriate concept for practical implementation. As an example, TBRS can be applied proactively by each CF while the CFs are still negotiating the channel allocation using SOCCER.

5.10   Concluding Remarks and Further References

Local area network deployment is expected to become increasingly important to provide mobile broadband because of the imminent spectrum scarcity. Reducing the cell size is one of the key tried and proved techniques to increase the area spectral efficiency. The importance of reduced cell sizes is likely to grow even more in the near future. Since femtocells are a relatively new frontier in cell granularity, the propagation and interference conditions can differ very much from those in larger cells.

Typically, in the initial femtocell deployments, intra-tier interference management is not a must due to the sparsity. Nevertheless, as the deployment density will increase smart channel allocation will be needed to guarantee a minimum quality for each femtocell, especially due to unfavorable topologies created by CSG deployment.

Because of the massive number of cells to be deployed, and the ad-hoc (potentially end-user) installation, the traditional approach to frequency planning and optimization will fail to be cost effective or even possible. Instead, self-organization of the channel allocation should be achieved with distributed dynamic channel selection algorithms, leading to the concept of CFs.

In this chapter we outline an analysis of the problem from different angles. From a single link perspective, a link experiencing low SINR is very sensitive to quality improvements but links with high SINR derive little benefit from interference reduction, especially due to MCS limitations. For this reason, a fine balance between interference avoidance and spectrum utilization needs to be found. The exact tradeoff in this analysis will differ according to the link level performance of a specific system.

When two or more links freely compete for spectrum access, the result is often destructive, and reuse 1 can be expected from independent adaptations. When the interference between the two links is correlated over time (a repeated game) there is incentive to achieve a better allocation. Even though the incentive exists, the different cells cannot certainly know about the incentive without signaling among cells. The third piece of analysis shows that even though a large number of cells may interact, the need of interference coordination is mostly characterized by the strength of interference and local interactions among neighbor cells. If the channel allocation

is done smartly one can avoid excessively sparse reuses while solving most of the interference conflicts. In the investigated scenarios the outage performance can be enhanced by adapting among reuses 1, 2, 3 and 4.

Three algorithms, with different levels of complexity, were briefly discussed. Their performances were evaluated using system-level simulations. These simulation results show that the distributed inter-cell interference management smartly performed among CFs is an effective way to share the spectrum among a very large number of radio links.

The reader interested in further details of these algorithms is referred to [5]. In particular, section 2.5 in [5] contains a quite extensive survey on spectrum sharing among CFs.

References

1. 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA): Further advancements for E-UTRA physical layer aspects. Technical report 36.814 v9.0.0, 2010
2. Ahmed, F., Tirkkonen, O., Peltomäki, M., Koljonen, J.-M., Yu, C.-H., Alava, M.: Distributed graph coloring for self-organization in LTE networks. JECE **2010**, 5:1–5:10 (2010)
3. Andrews, J.G., Claussen, H., Dohler, M., Rangan, S., Reed, M.C.: Femtocells: past, present, and future. IEEE J. Sel. Areas Commun. **30**(3), 497–508 (2012)
4. Claussen, H., Ho, L.T.W., Samuel, L.G.: Financial analysis of a pico-cellular home network deployment. In: IEEE ICC, Glasgow (2007)
5. da Costa, G.W.O.: Dynamic spectrum sharing among femtocells – coping with spectrum scarcity in 4G and beyond. PhD thesis, Radio Access Technology Section, Department of Electronic Systems, Aalborg University (2012)
6. da Costa, G.W.O., Cattoni, A.F., Kovacs, I.Z., Mogensen, P.E.: A fully distributed method for dynamic spectrum sharing in femtocells. In: 2012 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Paris, pp. 87–92 (2012)
7. Ellenbeck, J., Hartmann, C., Berlemann, L.: Decentralized inter-cell interference coordination by autonomous spectral reuse decisions. In: 14th European Wireless Conference (EW), Prague, pp. 1–7 (2008)
8. Etkin, R., Parekh, A., Tse, D.: Spectrum sharing for unlicensed bands. IEEE J. Sel. Areas Commun. **25**(3), 517–528 (2007)
9. Friedman, J.W., Mezzetti, C.: Learning in games by random sampling. J. Econ. Theory **98**(1), 55–84 (2001)
10. Fudenberg, D., Tirole, J.: Game Theory. MIT, Cambridge (1991)
11. Garcia, L.G.U., Pedersen, K.I., Mogensen, P.E.: Autonomous component carrier selection: interference management in local area environments for LTE-advanced. IEEE Commun. Mag. **47**(9), 110–116 (2009)
12. Garcia, L.G.U., da Costa, G.W.O., Cattoni, A.F., Pedersen, K.I., Mogensen, P.E.: Self-organizing coalitions for conflict evaluation and resolution in femtocells. In: 2010 IEEE Global Telecommunications Conference (GLOBECOM), Miami, pp. 1–6 (2010)
13. Garcia, L.G.U., Kovacs, I.Z., Pedersen, K.I., da Costa, G.W.O., Mogensen, P.E.: Autonomous component carrier selection for 4G femtocells – a fresh look at an old problem. IEEE J. Sel. Areas Commun. **30**(3), 525–537 (2012)
14. Goldsmith, A.: Wireless Communications, chapter 4, pp. 99–125. Cambridge University Press, Cambridge (2005)
15. Holma, H., Toskala, A.: LTE for UMTS – OFDMA and SC-FDMA based radio access. Wiley, Chichester (2009)

16. IST-4-027756 WINNER II, IST-WINNER D1.1.2: "WINNER II Channel Models", ver 1.1. Technical report, Winner II (2007)
17. Menon, R., MacKenzie, A.B., Buehrer, R.M., Reed, J.H.: A game-theoretic framework for interference avoidance in ad hoc networks. In: IEEE Global Telecommunications Conference (GLOBECOM'06), San Francisco, pp. 1–6, (2006)
18. Mogensen, P., Na, W., Kovacs, I.Z., Frederiksen, F., Pokhariyal, A., Pedersen, K.I., Kolding, T., Hugl, K., Kuusela, M.: LTE capacity compared to the Shannon bound. In: IEEE 65th Vehicular Technology Conference (VTC2007), Spring, Dublin, pp. 1234–1238 (2007)
19. Narayanan, L.: Channel assignment and graph multicoloring. In: Handbook of Wireless Networks and Mobile Computing, vol. 8, pp. 71–94. Wiley, New York (2002)
20. Pérez, D.L., et al.: OFDMA femtocells: a roadmap on interference avoidance. IEEE Commun. Mag. **47**(9), 41–48 (2009)
21. Saad, W., et al.: Coalitional game theory for communication networks: a tutorial. IEEE Signal Process. Mag. **26**(5), 77–99 (2009)
22. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
23. Suris, J.E., Dasilva, L.A., Han, Z., Mackenzie, A.B., Komali, R.S.: Asymptotic optimality for distributed spectrum sharing using bargaining solutions. IEEE Trans. Wirel. Commun. **8**(10), 5225–5237 (2009)

Chapter 6
Towards Cognitive Internet: An Evolutionary Vision

Fabrizio Granelli, Dmitry Kliazovich, and Neumar Malheiros

Abstract The requirement to support an always increasing number of networking technologies and services to cope with context uncertainties in heterogeneous network scenarios leads to an increase of operational and management complexity of the Internet. Autonomous communication protocol tuning is then crucial in defining and managing the performance of the Internet. This chapter presents an evolutionary roadmap of communication protocols towards cognitive Internet in which the introduction of self-aware adaptive techniques combined with reasoning and learning mechanisms aims to tackle inefficiency and guarantee satisfactory performance even in complex and dynamic scenarios.

In this survey, we overview and compare existing adaptive protocol stack solutions, review the principles of cross-layer design as well as the agent-based and AI based self-configuration solutions. The fundamental principles of cognitive protocols, such as adaptation, learning, and goal optimization, are presented along with implementation examples. Finally, the chapter discusses future research on the topic.

F. Granelli (✉)
DISI – University of Trento, Trento, Italy
e-mail: granelli@disi.unitn.it

D. Kliazovich
University of Luxembourg, Walferdange, Luxembourg
e-mail: dzmitry.kliazovich@uni.lu

N. Malheiros
University of Campinas, Campinas, Brazil
e-mail: ncm@ic.unicamp.br

6.1 Introduction

Communication networks play a central role in our daily life. Advanced wireless technologies and sophisticated network applications have changed the way we collaborate and share information. Today, the Internet represents the communication infrastructure of the modern society as well as a platform for the delivery of any kind of service (video on demand, e-commerce, social networks, etc.).

However, dealing with the complexity and dynamic of large scale networks is a challenging task. As the Internet represents a complex global interconnection of heterogenous end systems and networks, issues about its performance are rising and becoming increasingly relevant. The Internet architecture was not designed to support the Quality of Service (QoS) requirements of advanced multimedia applications and the varying channel conditions and mobility of wireless communication systems. The lack of adaptability and cross-layer mechanisms imposes limitations on internet protocols. In addition, current network management solutions face scalability and performance issues, as they are incapable of optimize resource utilization over long end-to-end paths and across heterogenous communication environments.

Academic and industry researches have been investigating how to provide adaptation capabilities to network protocols to optimize system-wide performance in a decentralized way. However, given the small and diverse time-scale of events and measurements in today's Internet, a specific requirement is related to autonomous operation (i.e. not requiring direct human intervention). Based on such principle, in this survey, we present a potential roadmap of the evolution toward a Cognitive Internet, in which distributed reasoning and learning mechanisms provide self-configuration and self-optimization capabilities to heterogenous networks elements.

This chapter is organized as follows. In the next section, we briefly describe some limitations of current Internet architecture and discuss the motivation for adaptive solutions. In Sect. 6.3, we overview well know techniques that enable design of adaptive protocol stack solutions. In Sect. 6.4, we focus on cognitive solutions to provide self-management functionalities. Then, we conclude the paper and discuss future research on the topic in Sect. 6.5.

6.2 Historical Perspective

6.2.1 Legacy TCP/IP

Core protocols of the TCP/IP stack were designed decades ago based on academic research requirements. The main design guidelines on which the Internet was built include layering, packet switching and the principle of keeping the complexity in the end systems at the edge of the network. These solutions are very elegant and effective, and they are still operational in today's Internet. However, a number of

inherited limitations prevent them to deliver the degree of efficiency, scalability and security dictated by current business and scientific requirements in networking.

The limitations of TCP/IP protocols have been described in the literature from several perspectives [18, 32, 37]. The main issues regard the lack of built-in security and support to mobility. A classic example is the poor performance of the traditional TCP congestion control mechanism in wireless environments. Although QoS solutions have been extensively investigated, there is no reliable and scalable approach to guarantee the stringent requirements of the advanced data-intensive network applications over the Internet. Moreover, the limited interaction among layers and lack of adaptation capabilities impact severely on service performance.

Several mechanisms have been proposed to overcome TCP/IP protocol drawbacks. As can be observed from the discussions in [31, 32, 37], the development of solutions for the next generation Internet addresses various issues such as service- and content-centric converged networks, new addressing schemes to support mobility, spectrum-efficient radio access, built-in security, and context-aware autonomic management capabilities. These solutions have been implemented if the form of extensions of the core architecture, affecting the transparency and simplicity of the original Internet design. As a consequence, ongoing discussion is aimed at the possibility or need for a new "clean" architecture for the Internet.

6.2.2 Motivation for Adaptation

Current Internet is built through the interconnection of different heterogeneous communication networks. This requires network administrators to deal with always increasing costs of operation and management. The ultimate solution would be to make management not necessary any more. In this context, the systems would be able to automatically manage themselves based on the high level administrative policies. This is called Autonomic Computing paradigm [10]. The self management of autonomic systems is based on the principles of context awareness, self-configuration, self-optimization, self-healing, and self-protection.

Obviously, the idea of autonomous control is far from original. The novelty in this paradigm is the holistic view of "autonomicity", and the focus of what is being automated. The goal is not automate system operation, but to automate its management functions. When the autonomic system is a communication network it is called an Autonomic Network [35]. The autonomic elements should be able to adapt themselves to constantly changing networks conditions in order to avoid performance degradation with minimum human intervention. The autonomic behavior must be guided by high level rules defined accordingly business and administrative policies.

The development of autonomic network management systems is a subject of considerable research and industrial interest [9, 36]. Network protocols can be adjusted during runtime using adaptive mechanisms, which make it a promising approach to deal with management complexity resulting from network heterogeneity and dynamics.

6.3 Adaptive TCP/IP: Enabling Technologies

In this section, we review the key technologies that enable the introduction of self-adaptation within the TCP/IP protocol stack. Those technologies are envisaged to build the basis on top of which to deploy cognitive TCP/IP solutions. The enabling technologies to support this evolution are:

- Cross-layer design, or cross-layering, providing suitable communication infrastructure for information and commands exchange among layers/protocols and network nodes;
- Distributed and agent-based solutions, enabling to re-allocate functionalities and features within the network;
- AI-based reasoning and learning, enabling the Internet to "think" and adapt;
- Architectures to support adaptive protocols, providing proper management environments.

As described in the following sections, cognitive networking employs cross-layer design, reasoning, and learning algorithms to provide system-wide network optimization through decentralized adaptation mechanisms.

6.3.1 Cross-Layer Design

The large variety of optimization solutions requiring information exchange between two or more layers of the protocol stack raises an important issue concerning implementation of different cross-layer solutions inside TCP/IP protocol reference model, their coexistence and interoperability, requiring the availability of a common cross-layer signaling model. This model defines the implementation principles for the protocol stack entities implementing cross-layer functionalities and provides a standardized way for ease of introduction of cross-layer mechanism inside the protocol stack. In the following, we review several cross-layer signaling paradigms that have been proposed by the research community.

6.3.1.1 Interlayer Signaling Pipe

One of the first approaches used for implementation of cross-layer signaling is revealed by Wang et al. [39] as interlayer signaling pipe, which allows propagation of signaling messages layer-to-layer along with packet data flow inside the protocol stack in bottom-up or top-down manner, as illustrated in Fig. 6.1. An important property of this signaling method is that signaling information propagates along with the data flow inside the protocol stack and can be associated with a particular packet incoming or outgoing from the protocol stack.

Fig. 6.1 Interlayer signaling pipe
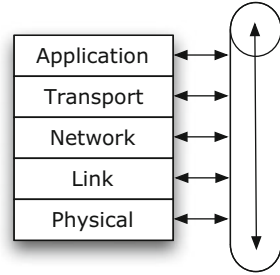
Two methods are considered for encapsulation of signaling information and its propagation along the protocol stack from one layer to another: packet headers or packet structures. Packet headers can be used as interlayer message carriers. In this case, signaling information included into an optional portion of IPv6 header [7], follow packet processing path and can be accessed by any subsequent layer. One of the main shortcomings of packet headers is in the limitation of signaling to the direction of the packet flow, making it not suitable for cross-layer schemes which require instant communication with the layers located on the opposite direction. Another drawback of packet headers method is in the associated protocol stack processing overhead, which can be reduced with packet structures method.

With packet structures signaling information is inserted into a specific section of the packet structure. Whenever a packet is generated by the protocol stack or successfully received from the network interface, a corresponding packet structure is allocated. This structure includes all the packet related information such as protocol headers and application data as well as internal protocol stack information such as network interface id, socket descriptor, configuration parameters and other. Consequently, cross-layer signaling information added to the packet structure is fully consistent with packet header signaling method but with reduced processing. Moreover, employment of packet structures does not violate existing functionality of separate layers of the protocol stack. In case the cross-layer signaling is not implemented at a certain layer, this layer simply does not fill nor modify the corresponding parts of the packet structure and does not access cross-layer parameters provided by the other layers. Another advantage of packet structure method is that standardization is not required, since the implementation could vary between different solutions.

6.3.1.2   Direct Interlayer Communication

Direct Interlayer Communication (Fig. 6.2) proposed in [39] aims at improvement of interlayer signaling pipe method by introducing signaling shortcuts performed out of band. In this way, the proposed Cross-Layer Signaling Shortcuts (CLASS) approach allows non-neighboring layers of the protocol stack to exchange messages, without processing at every adjacent layer, thus allowing fast signaling information delivery

Fig. 6.2 Direct interlayer
communication
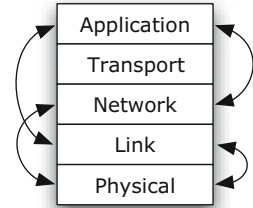
to the destination layer. Along with reduced protocol stack processing overhead,
CLASS messages are not related to data packets and thus the approach can be used
for bidirectional signaling. Nevertheless, the absence of this association is twofold
since many cross-layer optimization approaches operate on per-packet basis, i.e.
delivering cross-layer information associated with a specific packet traveling inside
the protocol stack.

One of the core signaling protocols considered in direct interlayer communi-
cation is Internet Control Message Protocol (ICMP) [33]. Generation of ICMP
messages is not constrained by a specific protocol layer and can be performed at any
layer of the protocol stack. However, signaling with ICMP messages involves oper-
ation with heavy protocol headers (IP and ICMP), checksum calculation, and other
procedures which increase processing overhead. This motivates a "lightweight" ver-
sion of signaling protocol CLASS which uses only destination layer identification,
type of event, and related to the event data fields. However, despite the advantages of
direct communication between protocol layers and standardized way of signaling,
ICMP-based approach is mostly limited by request-response action – while more
complicated event-based signaling should be adapted. To this aim, a mechanism
which uses callback functions can be employed. This mechanism allows a given
protocol layer to register a specific procedure (callback function) with another
protocol layer, whose execution is triggered by a specific event at that layer.

6.3.1.3 Central Cross-Layer Plane

Implemented in parallel to the protocol stack, the Central Cross-Layer Plane is
probably the most widely known cross-layer signaling architecture. In [4], the
authors propose a shared database that can be accessed by all layers for obtaining
parameters provided by other layers and providing the values of their internal
parameters to other layers, as illustrated in Fig. 6.3. This database is an example
of passive Central Cross-Layer Plane design: it assists in information exchange
between layers but does not implement any active control functions such as tuning
internal parameters of the protocol layers.

Similar approach is presented by the authors of [12], which introduces a central
cross-layer plane called Cross-layer Server able to communicate with protocols
at different layers by means of Clients. This interface is bidirectional, allowing
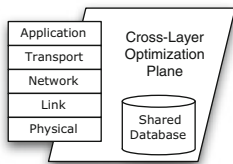
Fig. 6.3 Central cross-layer plane

Fig. 6.4 Network-wide cross-layer signaling

Cross-layer server to perform active optimization controlling internal to the layer parameters.

6.3.1.4 Network-Wide Cross-Layer Signaling

Most of the above proposals aim at defining cross-layer signaling between different layers belonging to the protocol stack of a single node. However, several optimization proposals exist which perform cross-layer optimization based on the information obtained at different protocol layers of distributed network nodes. This corresponds to network-wide propagation of cross-layer signaling information, which adds another degree of freedom in how cross-layer signaling can be performed, as illustrated in Fig. 6.4.

Among the methods overviewed above, packet headers and ICMP messages can be considered as good candidates. Their advantages, underlined in the single-node protocol stack scenario, become more significant for network-wide communication. For example, the way of encapsulating cross-layer signaling data into optional fields of the protocol headers almost does not produce any additional overhead and keeps an association of signaling information with a specific packet. However, this method limits propagation of signaling information to packet paths in the network. For that reason, it is desirable to combine packet headers signaling with ICMP messages, which are well suited for explicit communication between network nodes.

One of the early examples of cross-network cross-layering is the Explicit Congestion Notification (ECN) presented in [34]. It realizes in-band signaling approach by marking in-transit TCP data packet with congestion notification bit.

However, due to the limitation of signaling propagation to the packet paths this notification need to propagate to the receiver first, which echoes it back in the TCP ACK packet outgoing to the sender node. This unnecessary signaling loop can be avoided with explicit ICMP packets signaling. However, it requires traffic generation capabilities form network routers and it consume bandwidth resources.

An example of adaptation of Central Cross-Layer Plane-like architecture to the cross-network cross-layer signaling is presented in [20]. The chapter suggests the use of a network service which collects parameters related the wireless channel located at the link and physical layers, and then provides them to adaptive mobile applications.

A unique combination of local and network-wide cross-layer signaling approaches called Cross-Talk is presented in [40]. CrossTalk architecture consists of two cross-layer optimization planes. One is responsible for organization of cross-layer information exchange between protocol layers of the local protocol stack and their coordination. Another plane is responsible for network-wide coordination: it aggregates cross-layer information provided by the local plane and serves as an interface for cross-layer signaling over the network. Most of the signaling is performed in-band using packet headers method, making it accessible not only at the end host but at the network routers as well. Cross-layer information received from the network is aggregated and then can be considered for optimization of local protocol stack operation based on the global network conditions.

Main problems associated to deployment of cross-layer signaling over the network, also pointed in [21], include security issues, problems with non-conformant routers, and processing efficiency. Security considerations require the design of proper protective mechanism avoiding protocol attacks attempted by non-friendly network nodes by providing incorrect cross-layer information in order to trigger certain behavior. The second problem addresses misbehavior of network routers. It is pointed out that, in 70 % of the cases, IP packets with unknown options are dropped in the network or by the receiver protocol stack. Finally, the problem with processing efficiency is related to the additional costs of the routers? hardware associated with cross-layer information processing. While it is not an issue for the low-speed links, it becomes relevant for high speeds where most of the routers perform simple decrement of the TTL field in order to maintain high packet processing speed.

6.3.2   Distributed and Agent-Based Solutions

The possibility to abstract "atomic" functions from a specific protocol layer and execute them in the network is the base for distributed protocol stacks architectures (see Fig. 6.5). The design process is composed of the following procedures: abstraction, detachment, connection, and execution.
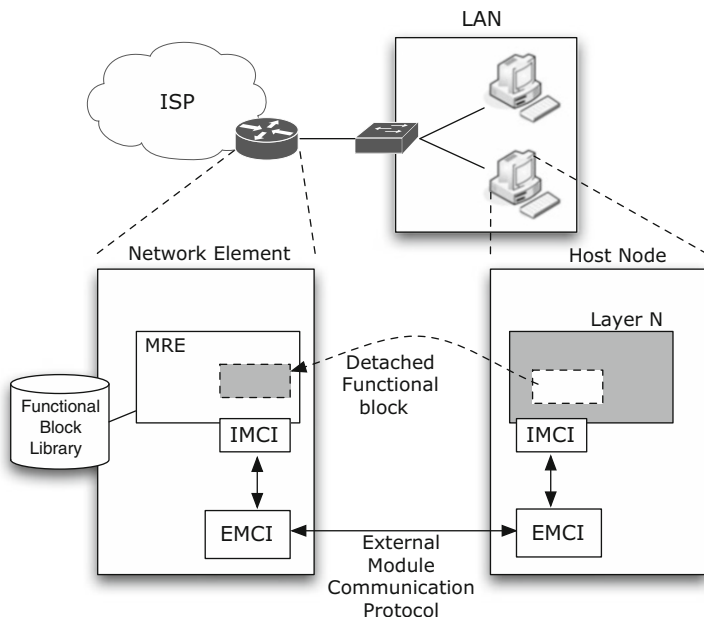
Fig. 6.5 Distributed protocol stack architecture

6.3.2.1 Abstraction

Before a specific function or a set of functions of the protocol stack can be distributed over the network, they should be abstracted and detached from the protocol stack of the host node.

Identification of the functions to be abstracted depends on the optimization goal and is performed on a case-by-case basis. However, as a general recommendation, an abstraction should be performed with non-time critical functions which work on packet basis and do not require continuous access to the internal kernel structures. Ideally, abstracted functions should fit into a single functional block which operates at a packet flow basis and requires minimum or no input from the host protocol stack. The output of the abstracted functional block should be applied to the packet flow (for example, controlling a single bit in a packet header), trying to avoid the requirement for direct communication with the host protocol stack.

Examples of protocol stack functional blocks that could be easily abstracted include TCP ACK generation module, header compression, IP security related functionalities, congestion related packet drop notification, advertise window adjustment in TCP, and many others.

6.3.2.2 Detachment

Once the identified set of functions is abstracted as a standalone functional block within a protocol layer, it can be detached and moved into the network. This procedure requires a certain level of cooperation from network elements (routers, switches, or gateways). In particular, network element can be considered "friendly" to the proposed Distributed Protocol Stack if they provide an environment able to support execution of the detached functional blocks – the Module Running Environment (MRE) – as an extension of their protocol stack.

MRE provides universal ways for registration and execution of different functional blocks. For example, it may provide a set of standard API functions which can be used by the host node to first transfer the abstracted functional block realized in the set of instructions understood by MRE (module description script language), and then register and run the transferred module.

Alternatively, avoiding the need for module transfer and registration procedures, functional blocks could be chosen from functional block library implemented at the network element. Execution of such blocks at the network element could be controlled by the host node or configured by network operator.

6.3.2.3 Communication

Communication between the detached functional block with the host protocol stack is performed using a Module Connection Interface (MCI), which is designed to provide communication between the detached functional block and the host protocol stack.

MCI is composed of two components:

- Internal Module Connection Interface (IMCI) connects the detached functional block with MRE at the network element side, while at the host node it provides communication interface with the protocol layer the functional block has been detached.
- External Module Connection Interface (EMCI) component provides communication between the detached functional block and the host protocol stack across the network with the use of External Module Communication Protocol (EMCP).

The main idea behind MCI separation into internal and external parts is designed for the purpose of module communication overhead reduction. In particular, EMCI components could be implemented at the lower layers or the protocol stack, leading to fewer header overhead and faster processing. The communication with the IMCI located at the protocol layer where the detachment was performed is performed locally within the protocol stack and it thus does not consume network resources.

6.3.2.4   Execution

Execution of the detached functional block can be triggered by the host node using MRE module installation primitives, or can be configured and running by the base station – without requiring interaction with the host node. In the latter case, the base station is responsible for notifying its clients with the information related to the list of functional blocks available.

Nevertheless, it is also important to consider the case of operating with clients which are unaware of functional blocks running at the base station or clients which do not support such operation. Operation of the detached functional blocks should be performed in a transparent way, causing no communication performance degradation.

6.3.3   AI-Based Reasoning and Learning

A set of solutions existing in the literature focus on the complexity of the optimization task, by proposing solutions based on Artificial Intelligence (AI). In this framework, emphasis is on the reasoning and learning processes of a cognitive network, i.e. on identifying suitable algorithms to understand the relationships among the network parameters with minimal a-priori knowledge. Such approaches can be based on fuzzy logic, reinforcement learning, or go beyond traditional AI towards bio-inspired operation [11]. The following paragraphs provide a brief review of AI-based solutions. For a more comprehensive and detailed analysis, the reader should refer to [13].

Expert systems, i.e., systems aiming to store human experts' knowledge in a specific field, represent a useful framework 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.

Heuristic optimization algorithms, like simulated annealing, genetic algorithms or swarm intelligence, are often used to automatically identify optimal solutions, and could be employed as alternative reasoning methods. However, such techniques should be preferred when the environment is well-known and the problem is centralized, rather than in distributed scenarios like the Internet.

Neural networks are often considered as a standard artificial intelligence technique and, thus far, 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, i.e. the existing relationships 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 networks 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. Similar issues apply to Markov random fields and Markov logic networks, even tough 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. An example of the usage of FCMs in networking is presented in [14]. Updating techniques are based on Hebbian learning, according to which connections between concepts that are activated together should be given more weight.

6.3.4   Architectures to Support Adaptive Protocols

Several approaches for adaptation through autonomic management have been proposed in the literature [30]. In [19], Jennings et al. proposed an autonomic network management architecture called FOCALE. This architecture makes use of information and ontological modeling to enable the system to learn and reason about itself and its environment. Such knowledge, embedded within system information models, is used by policy-based management systems to automatically configure network elements in response to changing environmental context. This realizes an autonomic control loop in which the system senses changes and enforces management actions accordingly.

The Autonomic Network Architecture (ANA) project [1] resulted in the design of a novel autonomic network architecture which provides generic networking abstractions and communication primitives to support network adaptability. As described by the authors, ANA was designed as a meta-architecture to support evolution and adaptation of novel networking mechanisms. ANA is a generic development framework and an execution environment for the development and testing of autonomic functionalities. In that work, the authors illustrate key flexibility features of ANA, such as, support to address-agnostic applications and node mobility within the so-called network compartments.

6.3.5   Discussion

In this section, we described key technologies that enable the design of adaptive mechanisms to optimize network performance. It is important to note that these

Table 6.1 Enabling technologies for adaptive protocols

Technology	Provided features
Cross-layer design	Interlayer communication to support improved performance analysis and decision-making
Distributed and agent-based solutions	Support to dynamic protocol stacks
AI-based algorithms	Enable the design of learning and reasoning mechanisms
Autonomic network architectures	Support to knowledge representation, signaling, and performance monitoring

technologies are complementary to each other and can be used together in the design of sophisticated self-management solutions. For example, a self-managed transport protocol may be designed by the combination of AI algorithms and cross-layer architecture. The AI algorithms can provide reasoning capabilities for optimal protocol tuning, while the cross-layer approach can provide effective system performance monitoring and analysis. Table 6.1 summarizes the main features provided by those technologies to support self-management capabilities.

6.4   The Evolution to Cognitive Protocols

Although researchers have well described requirements and architectures for the design of autonomic systems, concrete algorithms for the realization of self-management functionality remains a challenge. The design of decentralized adaptive mechanisms able to optimize system wide performance is a quite complex task.

Learning and reasoning techniques have emerged as a promising approach to provide adaptive capabilities to communication protocols. As far as we know, that was first introduced in the construct known as the Knowledge Plane, as proposed by Clark et al. [5]. They described a new goal for the next generation Internet: "the ability of the network to *know* what it is being asked to do, so that it can more and more take care of itself".

Another important proposal is the concept of Cognitive Networks [38]. As an evolution of the concept of cognitive radio [17, 29], the paradigm of cognitive networking combines cognitive algorithms, cooperative networking, and cross-layer design for the provisioning of real-time optimization of complex communication systems [22]. Indeed, cognitive radio focuses on the tuning of parameters at the physical and link layers to provide efficient spectrum sharing, while cognitive networking expands the dynamic tuning of parameters to a system-wide scale to improve overall network performance [15].

There are several proposals of cognitive network architectures in the literature [16, 24]. The Software Programmable Intelligent Network (SPIN) presented [26] merges concepts of IP, PSTN, cellular, and ad hoc networks for overcoming the fundamental limitations of IP networks. The SPIN architecture consists of three

planes interconnected by layer-2 transport infrastructure: the forwarding plane responsible for switching and monitoring, the control/management plane controlling the forwarding plane devices targeting flow optimization based on the received measurements, and the cognitive plane providing intelligence for and administration of the entire system.

Demestichas et al. present in [8] a platform (m@ANGEL) based on autonomic computing principles to provide seamless cognitive connectivity in heterogeneous wireless access networks. That work discusses business level issues and describes the architecture that provides management intelligence like context monitoring, description of profiles/agreements, resource brokerage, and configuration negotiation and implementation. The focus of m@ANGEL platform is exclusively devoted to bring cognitive functionalities into beyond-3G access networks. Most of the reconfiguration and cognitive functionalities are concentrated at the base stations. The structure of the access network consists of two planes: the infrastructure plane, which includes reconfigurable elements, and the management plane, composed of m@ANGEL entities responsible monitoring and control.

The Cognitive Complete Knowledge Network (CogNet) was proposed by Manoj et al. in [28]. It is a cross-layer approach aimed at extracting useful information from large amounts of network observations through inference algorithms and statistical learning techniques. The network state observations can be gathered through direct measurement or from peer nodes. The obtained knowledge enables optimal decisions in controlling network operation in order to improve the efficiency of resource management and the overall network performance.

In [25], Kousaridas et al. argue that next generation Internet architecture should be based on cognitive behavior by proposing a novel hierarchical feedback-control cycle in order to provide self-management capabilities. The proposed approach (called Self-NET) encompasses a hierarchical distribution of cognitive cycles in order to address self-organization and dynamic reconfiguration of network elements.

A cognitive network management architecture was proposed by Bouet et al. in [2]. The authors use software agents and artificial intelligence algorithms to build a distributed cognitive management framework (called CNM). Such framework includes communication, discovery and topology services, and uses a Fuzzy Logic-based inference system to support decision-making mechanisms. An important contribution of that work was to show the feasibility of embedding cognition into wireless networks. The paper discusses an implementation of the proposed cognitive architecture and its deployment within a heterogeneous wireless access network. The deployed solution was applied to two management functions, namely dynamic coverage control and capacity optimization.

In [27], Malheiros et al. present a feasible and effective solution for cognitive self-configuration of communication protocols. They propose a cognitive approach for dynamic reconfiguration of protocol parameters in order to avoid performance degradation as a consequence of changing network conditions. The proposed cognitive framework, called CogProt, provides runtime adjustment of protocol stack configuration parameters. The core of CogProt is a cross-layer cognitive plane, as illustrated in Fig. 6.6. CogProt periodically reconfigures the parameters

Fig. 6.6 The cognitive plane
architecture

Cross-Layer Cognitive Plane
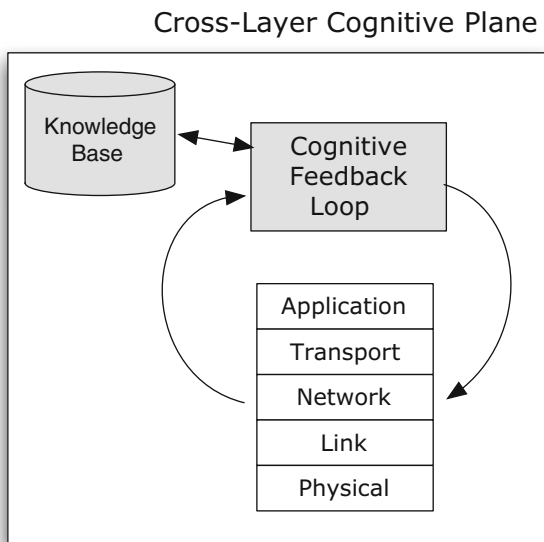
of interest based on acquired knowledge to improve system-wide performance. This dynamic reconfiguration process is implemented through a cognitive feedback which includes learning and reasoning mechanisms.

CogProt can be applied to a wide range of protocol parameters in different layers. As a proof of concept, the framework was illustrated for the cognitive configuration of TCP congestion window evolution as presented in [23]. The congestion window increase factor (α) controls the increase of TCP congestion window after each RTT period. Controlling the TCP window evolution allows the adjustment of network utilization, protocol fairness, and the level of network congestion. Higher α values are desirable in high bandwidth-delay network with low or moderate congestion levels, but should be avoided otherwise. However, there is no effective way for a network node to determine, in advance, available network bandwidth and the level of congestion at the end-to-end path between the sender and the receiver. Therefore, it is not possible to define an optimal value for α.

In that case study, CogProt was applied to the adaptation of the window increase factor which is adjusted in runtime based on the TCP goodput experienced in the immediate past. Both simulation and testbed experiments demonstrate that the proposed cognitive framework is able to improve average TCP performance under changing network conditions. The goal is to improve the performance of the congestion avoidance mechanism of standard TCP New Reno protocol for which the default value of the window increase factor is 1. We compared the TCP NewReno performance with fixed α values and dynamically adjusted α values (dynamically reconfigured by the cognitive mechanism). CogProt was allowed to vary the α parameter in the range [1, 5]. For fixed value of α, performance degrades under changing conditions. However, CogProt is able to keep performance close to the
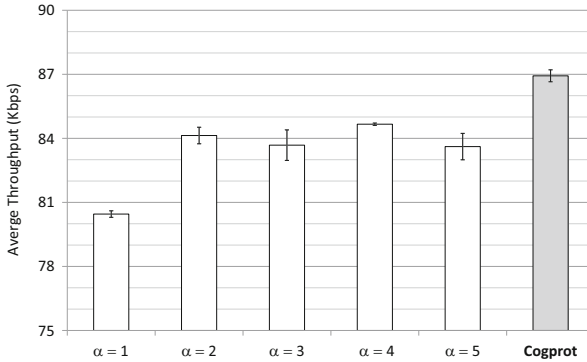
Fig. 6.7 Average TCP throughput for fixed α values and CogProt

optimal one for the scenarios with individual flows and it outperforms configurations with any fixed values of α, as shown in Fig. 6.7. These results demonstrate the benefit of CogProt dynamic adaptation capabilities to avoid performance degradation on varying network conditions.

In [3], CogProt was used to design a novel mechanism for rate adaptation in wireless networks. The distributed mechanism enhances network element with self-configuration functionality to dynamically adapt the MAC data rate. It is able to quickly react to changes on channel conditions in order to avoid performance degradation with fair resource sharing among nodes.

Another CogProt-based solution was proposed in [6]. In that work, the authors present a mechanism for cognitive optimization of multiple link layer parameters in wireless networks. That solution exploits the cognitive adaptation techniques to maintain link layer performance at the optimal level during runtime. Its effectiveness is demonstrated by tuning the parameters of the CSMA-CA protocol. Both simulation and experimental results confirm the benefits of the proposed cognitive adaptation strategy and the ability to maintain optimal configuration of link layer parameters, even under highly dynamic network conditions.

CogProt is a decentralized framework. Each network node is allowed to independently decide on its own protocol setup to best match the current network conditions. However, network nodes may share the knowledge and make collaborative reconfiguration decisions. To support them, the framework includes a centralized Cognitive Information Service (CIS) which fosters the exchange of cognitive information among the nodes belonging to the same network segment. However, when CIS service is not available the nodes can still share cognitive information in a completely decentralized way. CogProt provides both self-configuration during runtime and initial setup of protocol parameters. A cross-layer cognitive plane is in the core of self-configuration process at the network node level. It implements a control feedback loop requiring network nodes to build a knowledge base based on the evidenced average performance. Such performance information is used to periodically adjust the parameter of interest. CogProt does not require modifications

Table 6.2 Summary of cognitive approaches

Approach	Target layers	Decentralized
(SPIN) [26]	All	✓
m@ANGEL [8]	Lower	✓
CogNet [28]	All	×
Self-NET [25]	All	✓
CNM [2]	Lower	✓
CogProt [27]	All	✓

of standardized protocol operation and messaging. Thus, it is transparent to the rest of the network and can be deployed incrementally. The proposed framework can be used in the design of self-configuration mechanisms that are able to improve the average performance of network protocols with low degree of complexity and at low computational cost.

Table 6.2 summarizes the aforementioned cognitive approaches. We identify the target layers for which the solutions were designed and which of them are based on a decentralized approach.

6.5 Conclusion

In this work, we presented an overview of several approaches to cope with the complexity of managing dynamic and heterogeneous network environments. We emphasize the need for adaptive mechanism which should allow network elements to reconfigure themselves in order to avoid performance degradation in face of changing network conditions. Such self-management mechanisms must be decentralized and provide system-wide performance optimization.

In this scenario, cognitive techniques represent a promising approach to design the future Internet protocols. Several cognitive frameworks and mechanisms have been proposed to realize distributed self-management functionality. Nevertheless, there are challenges to be overcome in order to fulfill the requirements of scalable and decentralized adaptive network architectures. In this context, it is possible to identify the following issues as opportunities for future research on the topic:

- *Analytical Models*: Cognitive solutions aims at supporting dynamic adjustment of network systems to provide performance optimization. There is a lack of theoretical frameworks and analytical models which demonstrate whether the proposed solutions can efficiently converge to optimal operational points. Cognitive algorithms for learning and reasoning may have various control parameters. Such analytical models would also be very useful to determine optimal values for the control parameters themselves in specific network scenarios or applications.
- *Proactive Self-Management*: As far as we know, the proposed cognitive architectures and protocols provide network systems only with reactive reconfiguration functionalities. Network elements can adapt to changing environment conditions

in an attempt to maximize resource utilization, but only after performance degradation has already affected the system. It is worth to investigate novel self-management capabilities that are able to provide pro-active management. This way, network elements would be able to learn how to prevent performance degradation and security attacks.

- *Knowledge Representation*: In order to implement learning and reasoning algorithms in an efficient and scalable way, we need light and flexible knowledge information models. Such models must be independent of specific architectures and technologies, and facilitate cognitive information sharing among managed elements.

References

1. Bouabene, G., Jelger, C., Tschudin, C., Schmid, S., Keller, A., May, M.: The autonomic network architecture (ANA). IEEE J. Sel. Areas Commun. **28**(1), 4–14 (2010). doi:10.1109/JSAC.2010.100102
2. Bouet, M., Nguengang, G., Conan, V., Kousaridas, A., Spapis, P., Alonistioti, N.: Embedding cognition in wireless network management: an experimental perspective. IEEE Commun. Mag. **50**(12), 150–160 (2012). doi:10.1109/MCOM.2012.6384465
3. Chaves, L., Malheiros, N., Madeira, E., Garcia, I., Kliazovich, D.: A cognitive mechanism for rate adaptation in wireless networks. In: Strassner, J., Ghamri-Doudane, Y. (eds.) Modelling Autonomic Communications Environments. Lecture Notes in Computer Science, vol. 5844, pp. 58–71. Springer Berlin Heidelberg (2009). doi:10.1007/978-3-642-05006-0_5
4. Chen, K., Shah, S.H., Nahrstedt, K.: Cross-layer design for data accessibility in mobile ad hoc networks. Wirel. Pers. Commun. **21**(1), 49–76 (2002). doi:10.1023/A:1015509521662. http://dx.doi.org/10.1023/A%3A1015509521662
5. Clark, D.D., Partridge, C., Ramming, J.C., Wroclawski, J.T.: A knowledge plane for the internet. In: Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'03), Karlsruhe, pp. 3–10. ACM, New York (2003). doi:10.1145/863955.863957
6. de Oliveira Filho, J.L., Kliazovich, D., Granelli, F., Madeira, E., da Fonseca, N.: Cog-MAC: a cognitive link layer for wireless local area networks. Wirel. Netw. 1–11 (2013). doi:10.1007/s11276-012-0536-y. http://dx.doi.org/10.1007/s11276-012-0536-y
7. Deering, S., Hinden, R.: Internet Protocol, Version 6 (IPv6) Specification. IETF RFC 2460 (1998)
8. Demestichas, P., Stavroulaki, V., Boscovic, D., Lee, A., Strassner, J.: m@ANGEL: autonomic management platform for seamless cognitive connectivity to the mobile Internet. IEEE Commun. Mag. **44**(6), 118–127 (2006). doi:10.1109/MCOM.2006.1668430
9. Dobson, S., Denazis, S., Fernández, A., Gaïti, D., Gelenbe, E., Massacci, F., Nixon, P., Saffre, F., Schmidt, N., Zambonelli, F.: A survey of autonomic communications. ACM Trans. Auton. Adapt. Syst. **1**(2), 223–259 (2006). doi:10.1145/1186778.1186782
10. Dobson, S., Sterritt, R., Nixon, P., Hinchey, M.: Fulfilling the vision of autonomic computing. Computer **43**(1), 35–41 (2010). doi:10.1109/MC.2010.14
11. Dressler, F., Akan, O.B.: A survey on bio-inspired networking. Comput. Netw. **54**(6), 881–900 (2010). doi:http://dx.doi.org/10.1016/j.comnet.2009.10.024. http://www.sciencedirect.com/science/article/pii/S1389128610000241. New Network Paradigms
12. El Defrawy, K.M., El Zarki, M.S., Khairy, M.M.: Proposal for a cross-layer coordination framework for next generation wireless systems. In: Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing (IWCMC'06), Vancouver,

pp. 141–146. ACM, New York (2006). doi:10.1145/1143549.1143580. http://doi.acm.org/10.1145/1143549.1143580

13. Facchini, C.: Bridging the gap between theory and implementation in cognitive networks: developing reasoning in today's networks. Ph.D. thesis, University of Trento (2011)

14. Facchini, C., Holland, O., Granelli, F., da Fonseca, N.L., Aghvami, H.: Dynamic green self-configuration of 3G base stations using fuzzy cognitive maps. Comput. Netw. **57**(7), 1597–1610 (2013). doi:http://dx.doi.org/10.1016/j.comnet.2013.02.011. http://www.sciencedirect.com/science/article/pii/S1389128613000212

15. Fortuna, C., Mohorcic, M.: Trends in the development of communication networks: cognitive networks. Comput. Netw. **53**(9), 1354–1376 (2009). doi:10.1016/j.comnet.2009.01.002. http://dx.doi.org/10.1016/j.comnet.2009.01.002

16. Georgakopoulos, A., Tsagkaris, K., Karvounas, D., Vlacheas, P., Demestichas, P.: Cognitive networks for future internet: status and emerging challenges. IEEE Veh. Technol. Mag. **7**(3), 48–56 (2012). doi:10.1109/MVT.2012.2204548

17. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005). doi:10.1109/JSAC.2004.839380

18. Jain, R.: Internet 3.0: Ten problems with current internet architecture and solutions for the next generation. In: IEEE Military Communications Conference (MILCOM 2006), Washington, DC, pp. 1–9 (2006). doi:10.1109/MILCOM.2006.301995

19. Jennings, B., van der Meer, S., Balasubramaniam, S., Botvich, D., O Foghlu, M., Donnelly, W., Strassner, J.: Towards autonomic management of communications networks. IEEE Commun. Mag. **45**(10), 112–121 (2007). doi:10.1109/MCOM.2007.4342833

20. Kim, B.J.: A network service providing wireless channel information for adaptive mobile applications. I. proposal. In: IEEE International Conference on Communications (ICC), Helsinki, vol. 5, pp. 1345–1351 (2001). doi:10.1109/ICC.2001.937141

21. Kliazovich, D., Granelli, F., Redana, S., Riato, N.: Cross-layer error control optimization in 3G LTE. In: IEEE Global Telecommunications Conference (GLOBECOM'07), Washington, DC, pp. 2525–2529 (2007) doi:10.1109/GLOCOM.2007.480

22. Kliazovich, D., Granelli, F., Fonseca, N.L.S.D.: Architectures and cross-layer design for cognitive networks. In: Handbook on Sensor Networks, pp. 3–24. World Scientific, Singapore (2010)

23. Kliazovich, D., Malheiros, N., da Fonseca, N.L.S., Granelli, F., Madeira, E.: CogProt: a framework for cognitive configuration and optimization of communication protocols. In: Proceedings of the 2nd International Conference on Mobile Lightweight Wireless Systems (MOBILIGHT), Barcelona (2010)

24. Kliazovich, D., Granelli, F., da Fonseca, N.: Survey on signaling techniques for cognitive networks. In: IEEE 16th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Kyoto, pp. 5–10 (2011). doi:10.1109/CAMAD.2011.5941118

25. Kousaridas, A., Polychronopoulos, C., Alonistioti, N., Marikar, A., Mödeker, J., Mihailovic, A., Agapiou, G., Chochliouros, I., Heliotis, G.: Future internet elements: cognition and self-management design issues. In: Proceedings of the 2nd International Conference on Autonomic Computing and Communication Systems (Autonomics'08), Turin, pp. 13:1–13:6. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels (2008). http://dl.acm.org/citation.cfm?id=1487652.1487665

26. Lake, S.M.: Cognitive networking with software programmable intelligent networks for wireless and wireline critical communications. In: IEEE Military Communications Conference (MILCOM), Altantic City, vol. 3, pp. 1693–1699 (2005). doi:10.1109/MILCOM.2005.1605918

27. Malheiros, N., Kliazovich, D., Granelli, F., Madeira, E., da Fonseca, N.L.S.: A cognitive approach for cross-layer performance management. In: IEEE Global Telecommunications Conference (GLOBECOM 2010), Miami, pp. 1–5 (2010). doi:10.1109/GLOCOM.2010.5684313

28. Manoj, B.S., Rao, R.R., Zorzi, M.: CogNet: a cognitive complete knowledge network system. IEEE Wirel. Commun. **15**(6), 81–88 (2008). doi:10.1109/MWC.2008.4749751

29. Mitola, J., Maguire G.Q., J.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999). doi:10.1109/98.788210

30. Movahedi, Z., Ayari, M., Langar, R., Pujolle, G.: A survey of autonomic network architectures and evaluation criteria. IEEE Commun. Surv. Tutor. **14**(2), 464–490 (2012). doi:10.1109/SURV.2011.042711.00078

31. Pan, J., Paul, S., Jain, R.: A survey of the research on future internet architectures. IEEE Commun. Mag. **49**(7), 26–36 (2011). doi:10.1109/MCOM.2011.5936152

32. Paul, S., Pan, J., Jain, R.: Architectures for the future networks and the next generation internet: a survey. Comput. Commun. **34**(1), 2–42 (2011). doi:http://dx.doi.org/10.1016/j.comcom. 2010.08.001. http://www.sciencedirect.com/science/article/pii/S0140366410003580

33. Postel, J.: Internet control message protocol. IETF RFC 792 (1981)

34. Ramakrishnan, K., Floyd, S., Black, D.: The addition of explicit congestion notification (ECN) to IP. IETF RFC 3168 (2001)

35. Raymer, D., Meer, S.v.d., Strassner, J.: From autonomic computing to autonomic networking: an architectural perspective. In: Fifth IEEE Workshop on Engineering of Autonomic and Autonomous Systems (EASE 2008), Belfast, pp. 174–183 (2008). doi:10.1109/EASe.2008.26

36. Samaan, N., Karmouch, A.: Towards autonomic network management: an analysis of current and future research directions. IEEE Commun. Surv. Tutor. **11**(3), 22–36 (2009). doi:10.1109/SURV.2009.090303

37. Stuckmann, P., Zimmermann, R.: European research on future internet design. IEEE Wirel. Commun. **16**(5), 14–22 (2009). doi:10.1109/MWC.2009.5300298

38. Thomas, R., Friend, D., DaSilva, L., MacKenzie, A.: Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. IEEE Commun. Mag. **44**(12), 51–57 (2006). doi:10.1109/MCOM.2006.273099

39. Wang, Q., Abu-Rgheff, M.: Cross-layer signalling for next-generation wireless systems. In: IEEE Wireless Communications and Networking (WCNC 2003), New Orleans, vol. 2, pp. 1084–1089 (2003). doi:10.1109/WCNC.2003.1200522

40. Winter, R., Schiller, J.H., Nikaein, N., Bonnet, C.: CrossTalk: cross-layer decision support based on global knowledge. IEEE Commun. Mag. **44**(1), 93–99 (2006). doi:10.1109/MCOM.2006.1580938

Chapter 7
Automatic Best Wireless Network Selection Based on Key Performance Indicators

Stefano Boldrini, Maria-Gabriella Di Benedetto, Alessandro Tosti, and Jocelyn Fiorina

Abstract Introducing cognitive mechanisms at the application layer may lead to the possibility of an automatic selection of the wireless network that can guarantee best perceived experience by the final user. This chapter investigates this approach based on the concept of Quality of Experience (QoE), by introducing the use of application layer parameters, namely Key Performance Indicators (KPIs). KPIs are defined for different traffic types based on experimental data. A model for an application layer cognitive engine is presented, whose goal is to identify and select, based on KPIs, the best wireless network among available ones. An experimentation for the VoIP case, that foresees the use of the One-way end-to-end delay (OED) and the Mean Opinion Score (MOS) as KPIs is presented. This first implementation of the cognitive engine selects the network that, in that specific instant, offers the best QoE based on real captured data. To our knowledge, this is the first example of a cognitive engine that achieves best QoE in a context of heterogeneous wireless networks.

S. Boldrini (✉)
Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy

Department of Telecommunications, Supélec, Gif-sur-Yvette, France
e-mail: boldrini@newyork.ing.uniroma1.it

M.-G. Di Benedetto
Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Rome, Italy
e-mail: gaby@acts.ing.uniroma1.it

A. Tosti
Telecom Italia, Rome, Italy
e-mail: alessandro.tosti@telecomitalia.it

J. Fiorina
Department of Telecommunications, Supélec, Gif-sur-Yvette, France
e-mail: jocelyn.fiorina@supelec.fr

7.1 Introduction

Coexistence of different types of wireless networks is common experience. Widespread mobile devices use different technologies to communicate and exchange data. In most cases, when multiple networks are available, that may be based on either same or different technology, devices may choose the one to use and also possibly migrate from one network to a different one. This is for example the case when both cellular and one or more Wi-Fi networks are present.

Several investigations that defined algorithms for migration from a wireless network to another one of a different technology (the so-called *vertical handover* [1]) do exist. These "traditional" vertical handover algorithms are mainly based on physical or network layer parameters, or the combination of these two. In particular, Signal to Noise Ratio (SNR) and Received Signal Strength Indicator (RSSI) are the most studied and used parameters (even if usually linked to other network layer parameters) for the handover decision due to their simplicity [2]. This is "paid", however, with a lack of reliability in their real networks conditions representation.

Another important aspect is that, by considering lower layer parameters, the decision is taken with an eye on networks conditions; this is of course important, but only partial. In the process of network selection, more focus should be put, however, on final user experience, that can be better described and taken into account by the introduction of application layer parameters.

Moreover, network selection should be performed in an "intelligent" way, i.e. by adapting final decisions to a variety of factors such as the traffic type for which the connection needs to be established, networks current conditions and performance, as well as the used device performance.

This chapter aims at introducing the cognitive principle at the application layer by performing automatic best network selection based on "Key Performance Indicators". In other words, the final goal is the selection of the wireless network that can guarantee the best final user experience, thanks to the introduction of a *cognitive engine* that functions at the application layer.

To better understand and visualize this concept, a basic structure of the proposed model (deeply described in the following sections) is presented in Fig. 7.1.

The chapter is organized as follows. In Sect. 7.2 the concept of "Quality of Experience" is introduced and it is explained how it can be obtained considering "Key Performance Indicators"; focus is put on the case of Voice over IP traffic type. Section 7.3 introduces the *cognitive engine*, whose behaviour and functionalities are described in Sect. 7.4. Section 7.5 presents an experimentation of the presented cognitive engine module in the Voice over IP case, while Sect. 7.6 contains conclusions and future work.
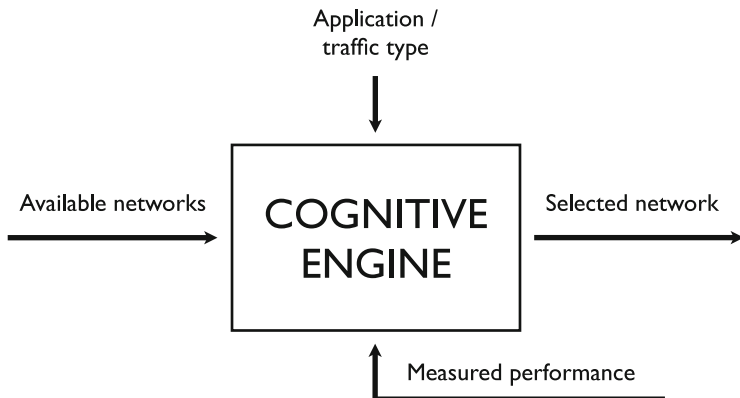
Fig. 7.1 Basic structure of the proposed cognitive engine

7.2 Quality of Experience and KPIs

Quality of Service (QoS) is nowadays a fundamental aspect that Internet Service Providers (ISPs) have to take into account in order to offer different services with different guaranteed qualities at different prices. Parameters that are traditionally considered for QoS belong to physical layer (SNR/RSSI) or network layer (delay, jitter, throughput and packet loss). These values are the ones from which a QoS profile or QoS classification is built on. In other words, these parameters determine a classification in different traffic classes, each one with a different quality.

From the final user point of view, these parameters are only values that characterize its communication. What the user is really interested on, however, is the final quality perceived and experienced. This aspect is the reason for moving on from Quality of Service to "Quality of Experience" (QoE) [3, 4]. For example, the delay a network presents is an important factor that has impact on user's QoE; anyway, delay itself, considered alone and not in the whole context, does not completely reflect the quality the user effectively experiences.

Since the goal of offering a certain level of quality must focus on the final user, the quality that is effectively experienced must be pursued. There is the need of parameters that are better able to represent the perceived quality: "Key Performance Indicators" (KPIs).

KPIs are application layer parameters and therefore are much closer to the truly experienced quality. Given that these reside on a higher layer of the OSI model, they include and take into account the previous mentioned parameters, but in a wider and more comprehensive context. In fact they are able to consider all lower layers parameters and "synthesize" them by giving them the appropriate "weight" (for example in a linear combination, as presented in Sect. 7.3) based on the considered traffic type. SNR, delay, jitter, throughput and packet loss are therefore not important by themselves, but as part of more general parameters that incorporate them. Thanks to a learning process, KPIs are able to include also delays introduced by particular

implementations of softwares and firmwares and specific behaviours of different devices using different telecom companies, aspects that are proved to be significant in the final user experience [5].

Until now, by our knowledge, application layer parameters have been introduced regarding minor aspects and in very specific cases [6, 7]. This chapter proposes to introduce an extensive use of these parameters for QoE evaluation.

KPIs can be defined for different traffic types, as for example voice communication, video and audio streaming, and web browsing. Each traffic type has its own peculiarities and "weaknesses", and therefore the attention on different aspects needs to be put based on the traffic type that is under consideration. For example, the delay a network presents is always important, but the impact it has on voice traffic type is considerably higher then in the case of web browsing traffic type; a similar thing can be said when considering jitter.

The identification and definition of the most suitable KPIs and their dependence on lower layers parameters for each traffic type can be done through an analysis of traffic data. Thanks to these data, the perceived quality can be correlated to the different layer parameters that result to be the most relevant (for the traffic type under consideration) and that therefore need to be considered for the KPIs definition. Traffic data used in this chapter for the definition of KPIs were provided by one of the major Italian telecom operator, that actively contributed in this work.

7.2.1   VoIP Case

This chapter focuses on "Voice over Internet Protocol" (VoIP) traffic. This traffic type was specifically investigated because it represents nowadays an increasing relevance in Internet traffic (shown by the high popularity of specific software applications and the services offered by ISPs) and can also be an interesting study-case to test the proposed approach.

Two KPIs were identified to be relevant for VoIP traffic type:

1. One-way end-to-end delay (OED);
2. Mean Opinion Score (MOS).

OED, as the name says, is the unidirectional delay that is encountered from the sending node to the receiving node. Its value is the sum of every delay contribution introduced by each network node passed through. An indication of unidirectional delay values related to the quality of the communication can be found in [8]. International Telecommunication Union (ITU) indicates two threshold values: if the one-way delay is below 150 ms, the quality is very good; if the one-way delay is above 400 ms, the quality is very poor.

Based on this indication and on the provided data, in this chapter the following association between delay threshold values and perceived quality was used:

- If OED $\leq$ 150 ms, the communication perceived quality is very good;
- If 150 ms $<$ OED $\leq$ 250 ms, the communication perceived quality is quite good;

Table 7.1 Association among MOS values, voice communication quality and perceived disturb

MOS	Communication quality	Disturb description
5	Very good	Not perceivable
4	Good	Slightly perceivable
3	Medium	Perceivable but not annoying
2	Poor	Annoying
1	Very poor	Very annoying

Table 7.2 Second model used for MOS estimation

MOS	Packet loss (%)	Jitter (ms)
5	0	0
4	≤ 3	≤ 75
3	≤ 5	≤ 125
2	≤ 10	≤ 125
1	> 10	> 125

- If $250\,\text{ms} < \text{OED} \leq 450\,\text{ms}$, the communication perceived quality is medium/poor;
- If $\text{OED} > 450\,\text{ms}$, the communication perceived quality is very poor.

MOS is a score that indicates the quality of a voice communication; it may vary in a range that goes from the minimum value of 1, that corresponds to a very poor quality, to the maximum value of 5, that corresponds to a very good quality [9]. It derives historically from the mean score assigned in tests with listeners in determined conditions. An association among MOS values, voice communication quality and perceived disturb can be found in Table 7.1.

In this chapter, two models for the MOS estimation were used. The first is described in [10] and can be expressed by the following equation:

$$\text{MOS} = 4 - 0.7 \cdot ln(\text{loss}) - 0.1 \cdot ln\left(\frac{\text{M} - \text{hsize}}{\text{drate}}\right),$$

where "loss" is the packet loss expressed in percentage, "M" is the IP packet dimension expressed in bytes, "hsize" is the IP packet header dimension expressed in bytes, and "drate" is the used codec datarate expressed in kilobytes per second (kB/s). This model is valid in IP networks, and consider 4 as MOS maximum value. Other more complex models for MOS estimation can be found in [11, 12].

The second model used for MOS estimation derives from the provided traffic data and is summarized in Table 7.2. In this case, differently from the first model used, jitter is taken into account. A MOS value is assigned to a voice communication if it respects both the corresponding values imposed for packet loss and jitter (see Table 7.2). As an example, if from a determined number of sent packets it is obtained a packet loss of 2 % and a mean jitter of 100 ms, a MOS value of 3 is assigned. Note that with this model only discrete MOS values are assigned, and that a MOS value of 5 is theoretically possible, even if practically quite impossible to obtain.

7.3   Cognitive Engine

This chapter proposes the introduction of a module called "cognitive engine", that can be implemented and installed in mobile devices. The final goal of the cognitive engine is to identify and select the wireless network, among the available ones, that permits to offer the best QoE for the final user. The network selection is based on KPIs, and for this reason is valid for a specific type of traffic.

Since the decision must be taken considering all the KPIs defined for the selected traffic type, a rule for the final selection that includes all of them must be defined. In this chapter, the definition of a *cost function* is proposed. In particular, a simple linear combination of KPIs is proposed as cost function. Given an application, i.e. a traffic type, a wireless network and related KPI values, the cost is therefore expressed by the following equation:

$$c\left(\text{KPI}_1, \ldots, \text{KPI}_N\right) = \sum_{i=1}^{N} g_i \, \text{KPI}_i \, ,$$

where c is the final cost value of the network, N is the number of KPIs considered for the actual traffic type, and g_i is the gain for the ith KPI.

It must be noted that each KPI presents a different gain g, i.e. has a different weight on the final decision. The gain values, that is to say how much a KPI is important for the cost within a specific traffic type, are determined by experimental data (together with the KPIs definition). However, the system presents a high flexibility. In fact the gain values can be updated and adjusted thanks to a learning process in order to refine the final selection based on the device specific behaviour (its firmware and software implementations, as better explained in Sect. 7.2).

Obviously, the goal is to obtain the lowest possible cost. This means that the selected wireless network is the one that presents the lowest cost. In this way, a soft decision is taken. However, for specific application or traffic types, it might be necessary to slightly modify this by introducing a hard decision rule. For example, in specific cases a KPI can be much more important than the others for QoE, and this can condition the final network selection.

7.4   Model Structure

The cognitive engine is designed to be an intermediate layer of the system, considering an Open Systems Interconnection (OSI) protocol stack model. It is located right under the application layer, so that it can communicate directly with the applications that are running in the device. It is also in direct communication with the operating system (OS) of the device, in order to obtain information about the available wireless networks and the connection status of the current network in terms of lower layers parameters (SNR, delay, jitter, throughput, packet loss
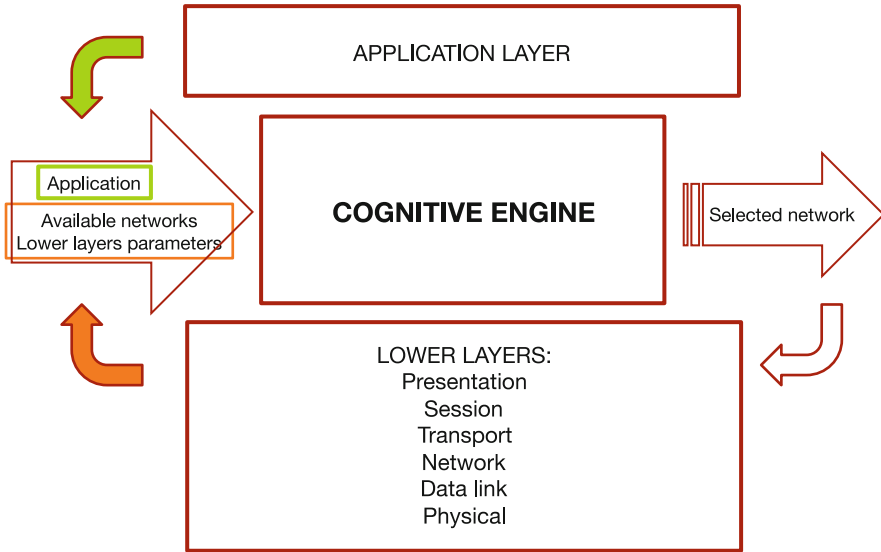
Fig. 7.2 The cognitive engine as intermediate layer and its location in the OSI system model

and every other parameter eventually necessary for the KPIs computation). This model structure is shown in Fig. 7.2. Note that the cognitive engine is thought to be used for the wireless network selection, so all data not implied in the selection process can skip the transition through the cognitive engine and can directly pass from application to presentation layer.

The inputs of the cognitive engine are the following:

- The application that needs a connection (from the application layer);
- The available wireless networks (from the OS);
- Lower layer parameters, eventually necessary for the KPIs computation (from the OS).

The output of the cognitive engine consists in the selected wireless network, that scores the best values of KPIs relevant for the running application (i.e. for the relative traffic type).

The functional behaviour of the cognitive engine can be outlined by the following logical steps:

- the application received as input is associated to one of the defined traffic types;
- once the traffic type is selected, the corresponding KPIs that need to be used and evaluated are identified;
- lower layers parameters that are needed for the KPIs evaluation are identified;
- for each available wireless network (whose list is received as input) lower layers parameters identified in the previous step are obtained:

- from memory (input in the cognitive engine), if a previous measurement step was carried out;
- by measuring; a "trial" connection is established if needed (this is the case, for example, of network layer parameters, that cannot be obtained otherwise);

- these parameters are used for KPIs evaluation, based on KPIs definitions and models; for each network there is, therefore, a set of KPI values;
- based on the KPI values, the cost function is computed for each network;
- the wireless network that presents the lowest cost is selected: it is the output of the cognitive engine.

Obviously, networks conditions may change: new wireless networks may be available, others may cease to be available (especially considering that the cognitive engine is thought to be implemented in mobile devices), and furthermore networks conditions may vary, so that the resulting KPIs may become significantly different from the values previously considered. Given this context, a periodical update must be performed in order to guarantee the choice of the best network under variable conditions. For this reason, the cognitive engine periodically updates the list of available networks and corresponding KPIs by periodically repeating the steps presented above. This means that measures with the currently selected network are periodically performed and KPIs values updated; "trial" connections are again established for the other networks in order to have also their KPIs to be compared to the other values and, if convenient, a different network selection can be done. The frequency of this periodical update must be discussed separately since it involves many different aspects.

Moreover, the cognitive engine must incorporate a learning mechanism. Since each different device behaviour may introduce different delays and performance modifications that can significantly affect QoE [5], one of the cognitive engine task must be to "learn" from the device behaviour, to adapt to it, and to react as a consequence. To react means to consider the performance of the devices, i.e. to include, for example, the delay introduced by the specific implementation of the Internet browser or the VoIP application in the device where the cognitive engine is running. These behaviours cannot be known a priori, and for this reason a learning step is required.

Also the different gains for the different KPIs used for the cost function evaluation can be updated, adapted and modified. As a drawback, this learning phase can take some time, but it can be done in background as a "refining" process for the network selection. But the big advantage is a very refined selection method that completely takes into account *all* the aspects involved in the quality perceived by the final user.

A final consideration must be done on the so called "ping-pong effect". After an update, if the selected network is different from the one selected in a previous stage, the device OS must connect to the new network in order to perform the best QoE. However, given the variability of the channels corresponding to the different networks (and especially if the cost variation is quite low), a same previous network could be selected in a following update. If the network change was

performed, then at a next stage another change will be necessary, causing continuous and unnecessary network choice fluctuations, with the consequence of a waste of resources in terms of time and energy consumption spent for performing the changes. For this reason, in order to avoid this "ping-pong effect", a latency on the decision or a hysteresis with threshold must be applied before effectively deciding for a network change. This also means that a series of frequent updates must be performed before proceeding with a change in the selected network.

7.5 Experimentation

7.5.1 Experimental Set-Up

A first experimentation was carried out considering the VoIP case. The parameters needed for the computation of OED and MOS (the two KPIs identified for VoIP traffic type, as presented in Sect. 7.2) are therefore the following:

- end-to-end delay;
- packet loss;
- IP packet dimension and its header dimension;
- used codec datarate;
- jitter.

Three of the above parameters (delay, packet loss and jitter) were obtained thanks to the use of the *ping* utility; the remaining (packet and header dimensions and datarate) were set as *ping* or KPIs inputs.

Packets sent with *ping* were sent from a computer towards a website server; this was chosen in order to always guarantee a minimum number of hops passed through and have therefore a realistic situation where the two end devices are connected through a certain number of intermediate nodes. In this case, there were always at least 15 hops from every location where the *ping* capture was carried out to the website server. For each capture, 50 packets of 64 bytes (dimension of IP packets) were sent. The obtained values are the result of the average of the 50 packets sent (and received back). Captures were taken at different times of the day during 10 days.

Five different wireless networks were used for the captures: three Wi-Fi networks and two different connections to the cellular network. These networks are located in different places in Rome, Italy; in every place, though, the minimum number of hops was respected. Although they are not present in the same place, they are an example of different wireless networks that can be effectively found in a same place and among which the device must choose. (They were chosen in different places due to captures bonds related to timing and capture device availability.) The considered codec for VoIP communication is G.729 (CS-ACELP, conjugate-structure algebraic-code-excited linear prediction), that provides a datarate of 8 kB/s.
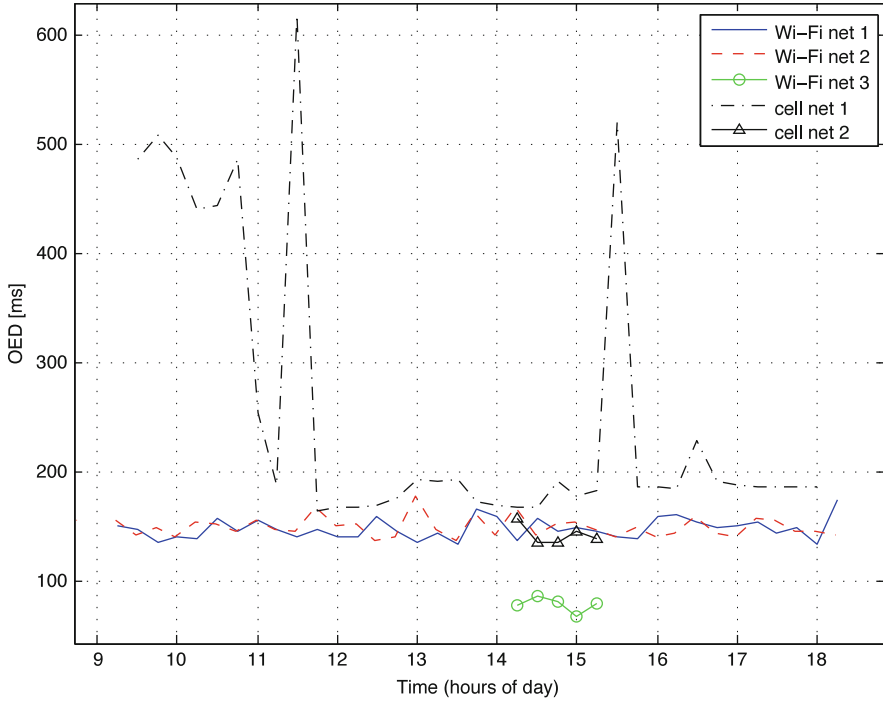
Fig. 7.3 OED values obtained at different times of the day with different wireless networks

Values of the parameters obtained through these *ping* captures, together with the set values, were used to compute the KPIs for VoIP, and the final KPI values were then stored with the association to the time of the day when the capture was taken. In this first experimentation, a granularity of 15 min was considered, i.e. a capture was performed every 15 min during the central hours of the days.

7.5.2 Experimental Data

OED values obtained are shown in Fig. 7.3. For two of the networks (Wi-Fi network 3 and cellular network 2) data are available only for limited times of the day (between 14.15 and 15.15). Wi-Fi networks present in general lower delay (OED values are lower). Moreover cellular networks (in particular cellular network 1) show much more delay variability.

MOS values obtained are shown in Fig. 7.4 (first model used) and in Fig. 7.5 (second model used). It can be easily seen that the second model presents only discrete values. For both models the maximum value is limited to 4. It must be noted that when jitter is considered for MOS evaluation, i.e. in the second model
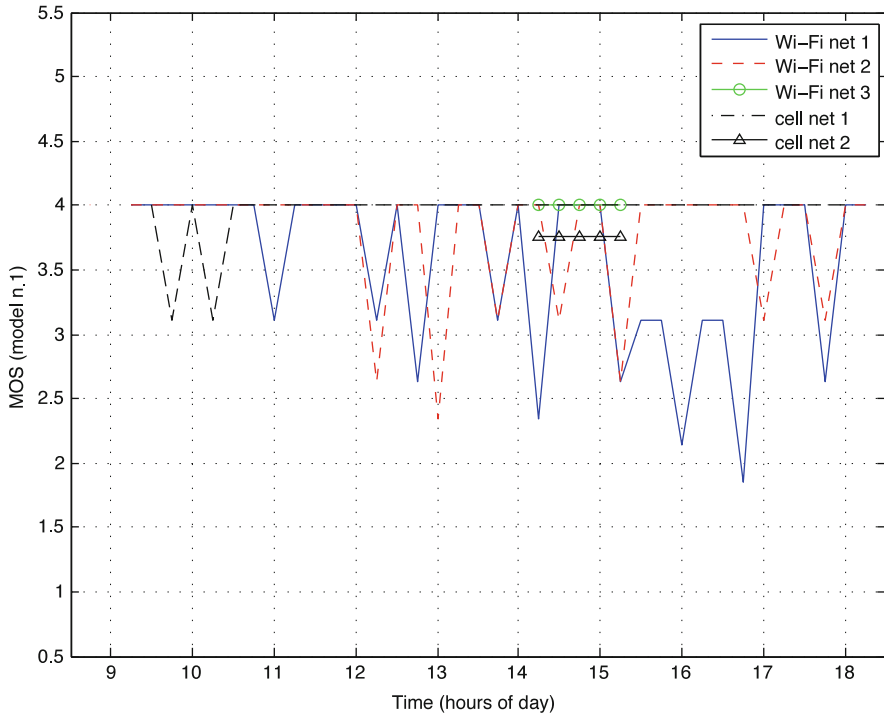
Fig. 7.4 MOS values obtained using the first model at different times of the day with different wireless networks

used, cellular network 1 shows much lower MOS values in moments of the day when there is more variability in the delay.

In a first implementation of the cognitive engine, used for testing the described system, memorized data were used in order to select the wireless network that offers the best QoE for VoIP traffic; all presented networks were therefore thought to be available in the same place. Collected KPI data were normalized and used for network selection. Gain values chosen are 0.7 for OED and 0.3 for MOS, that is to say that end-to-end delay is considered to weight 70 % on the QoE in a VoIP communication, and MOS is considered to weight the remaining 30 %. These gain values can be, as explained before, adjusted and updated. Repeating the selection process at different times of the day gave as a result different networks, according to the memorized KPI values.

Cognitive engine must update the KPI values of the available networks before making the selection, in order to have the estimate of the experienced quality the more realistic as possible. However, having a database with KPI memorized data of the network that were already "seen" in the past permits to have a first estimate in case the update process is not possible before the application requires a connection establishment (for example because there is no time to complete the update process
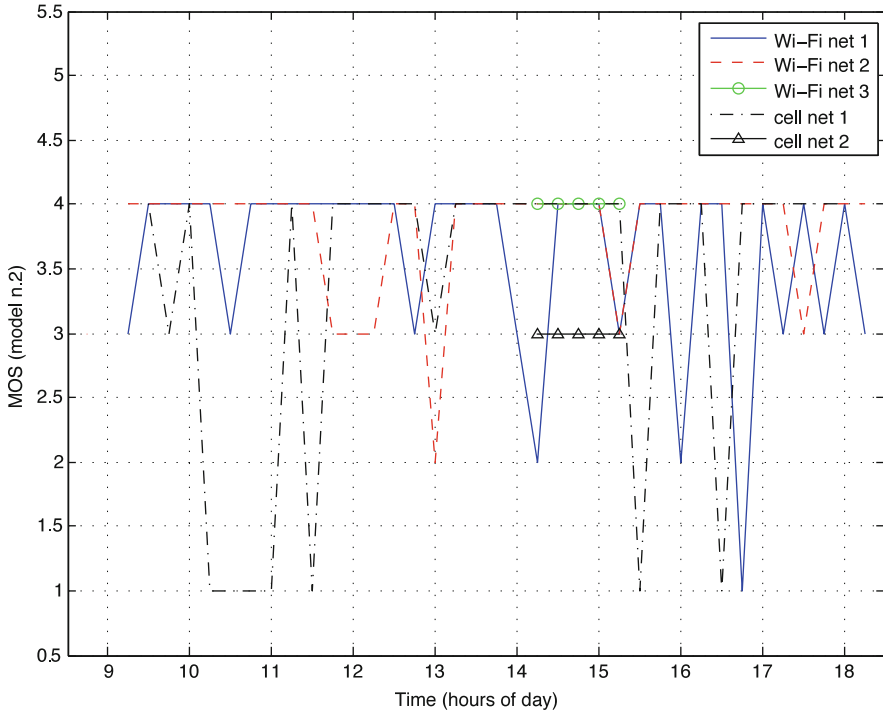
Fig. 7.5 MOS values obtained using the second model at different times of the day with different wireless networks

before the connection starts). This estimate can be "rough" if it is based on few data, but it is at least a first basis on which the decision can be taken; moreover, as soon as possible, new values can be collected, data can be updated and the estimate can be therefore refined.

A consideration should also be done on the initial transitional period. In fact, when a new network, that was never "seen" before, becomes available, there is no stored data related to it. Until a new KPI update is performed, in order to have data also of this network, it is not selected even if it can present performance able to permit the best experience for the final user. In the period before the new update it is therefore present a transitory, where the best QoE is not fully guaranteed due to the lack of data.

7.6 Conclusions and Future Work

In this chapter, a cognitive architecture was introduced at the application layer: the wireless network that can guarantee the best experience to the final user was automatically selected thanks to the introduction and use of application layer

parameters, i.e. Key Performance Indicators. Quality of Experience was introduced and KPIs were defined for different traffic types based on experimental data. The model of a cognitive engine was presented, whose goal is to identify and select, based on KPIs, the best wireless network among the available ones. An experimentation was then carried out considering the VoIP case, with OED and MOS as KPIs.

From our knowledge, this is the first case in which application layer parameters are used in an extensive way, and the first example of cognitive engine with the goal of achievement the best QoE in a context of heterogeneous wireless networks.

The presented system presents high flexibility, since it can be applied in a general context, with different wireless technologies and with different types of traffic.

This cognitive engine model, that was tested in the VoIP case, should be tested with other traffic types, introducing the appropriate KPIs. Future work on this topic will also focus on the selection algorithm: the convergence time to the best network must be minimized, by taking into account the "multi-armed bandit problem", i.e. how often the measuring (update) step should be performed and when avoiding wasting resources for the update process. Moreover, the presence of multiple users should be considered, since it may affect and modify performance of the different networks.

References

1. Yan, X., Sekercioglu, Y.A., Narayanan, S.: A survey of vertical handover decision algorithms in fourth generation heterogeneous wireless networks. Comput. Netw. **54**, 1848–1863 (2010)
2. Ahmed, A., Boulahia, L., Gaiti, D.: Enabling vertical handover decisions in heterogeneous wireless networks: a state-of-the-art and a classification, IEEE Commun. Surv. Tutor. **16**(2), 776, 811 (2014)
3. Piamrat, K., Viho, C., Ksentini, A., Bonnin, J.-M.: Quality of experience measurements for video streaming over wireless networks. In: 6th International Conference on Information Technology: New Generations (ITNG'09), Las Vegas, Nevada, USA, 27–29 April 2009
4. Jelassi, S., Rubino, G., Melvin, H., Youssef, H., Pujolle, G.: Quality of experience of VoIP service: a survey of assessment approaches and open issues. IEEE Commun. Surv. Tutor. **14**(2), 491–513 (2012)
5. Huang, J., Xu, Q., Tiwana, B., Mao, Z.M., Zhang, M., Bahl, P.: Anatomizing application performance differences on smartphones. In: MobiSys, San Francisco (2010)
6. Wang, C., Lin, T., Chen, J.-L.: A cross-layer adaptive algorithm for multimedia QoS fairness in WLAN environments using neural networks. IET Commun. **1**(5), 858–865 (2007)
7. Si, P., Ji, H., Yu, F.R.: Optimal network selection in heterogeneous wireless multimedia networks. Wirel. Netw. **16**, 1277–1288 (2010). Springer
8. ITU-T G.114, One-way transmission time (2003)
9. ITU-T P.800, Methods for subjective determination of transmission quality (1996)

10. Yamamoto, L.A.R., Beerends, J.G.: Impact of network performance parameters on the end-to-end perceived speech quality. In: Expert ATM Traffic Symposium, Mykonos, Greece, September 1997
11. Ding, L., Goubran, R.A.: Speech quality prediction in VoIP using the extended E-model. In: GLOBECOM, San Francisco (2003)
12. Sun, L., Ifeachor, E.C.: Voice quality prediction models and their application in VoIP networks. IEEE Trans. Multimed. **8**(4), 809–820 (2006)

Chapter 8
Localization in Cognitive Radio Networks

Ioannis P. Chochliouros, Ioanna Papafili, George S. Agapiou,
Anastasia S. Spiliopoulou, Stelios Agapiou, Ronald Raulefs, and Siwei Zhang

Abstract Cognitive Radio Networks (CRNs) are emerging as a viable solution to solve spectrum shortage problems, with the view of the evolutionary progress towards realizing the network of the future. The present work discusses a general survey of several among the critical capabilities and/or features characterizing CRNs, in the context of actual European standardization efforts. The chapter outlines an overall and harmonized technical concept for future CR systems, especially by discussing several options affecting the future evolution of radio technologies and network architectures towards more flexible and reconfigurable CR systems, as the latter are expected to increase the efficiency of the overall spectrum usage by offering new sharing opportunities and thus to provide more flexibility to applications-services.

I.P. Chochliouros (✉) • I. Papafili • G.S. Agapiou • A.S. Spiliopoulou
Hellenic Telecommunications Organization S.A. (OTE), Maroussi, Greece
e-mail: ichochliouros@oteresearch.gr; iopapafi@oteresearch.gr; gagapiou@oteresearch.gr;
aspiliopoul@ote.gr

S. Agapiou
Department of Informatics and Telecommunications, National Kapodistrian University of Athens,
Athens, Greece
e-mail: sdi1000008@di.uoa.gr

R. Raulefs • S. Zhang
Institute of Communications and Navigation of the German Aerospace Center (DLR), Wessling,
Germany
e-mail: ronald.raulefs@dlr.de; siwei.zhang@dlr.de

8.1 Cognitive Radio: An "Enabler" for Future Evolution of Communications Systems

Cognitive Radio Networks (CRNs) are emerging as a viable solution to solve spectrum shortage problems, with the view of the evolutionary progress towards realizing the network of the future. The present work discusses a general survey of several among the critical capabilities and/or features characterizing CRNs, in the context of actual European standardization efforts. The paper outlines an overall and harmonized technical concept for future CR systems, especially by discussing several options affecting the future evolution of radio technologies and network architectures towards more flexible and reconfigurable CR systems, as the latter are expected to increase the efficiency of the overall spectrum usage by offering new sharing opportunities and thus to provide more flexibility to applications-services.

Cognitive Radio (CR) has first been identified as a favourite technology for high-end applications in the military and public safety domain. Following to this original concern, several research-oriented efforts have examined various options and/or potential alternatives for "inclusion" of CR-based concepts to the requirements of civil wide area (cellular) and short-range communication systems [1]. A variety of market "actors" with numerous representatives from the industry, the regulation and the academia have realised applied research around the essential CR-based context, that is the option for realizing an enormously increased level of spectral efficiency, with "enhanced" overall system capacity exploitation. This has been performed by considering: (i) The dual exploitation/utilization of spectrum by applying methods for opportunistic spectrum usage, and; (ii) the option for a part of an equipment or device – or Mobile Device (MD) – to be (entirely or in part) conscious of its (heterogeneous) broader "context" (i.e., radio context, application context, etc.) and to adapt, vigorously, its operational parameters so that to achieve an "optimum" of its functioning objectives (e.g., for a MD to become aware of any adjacent Radio Access Technologies (RATs) and to choose the most appropriate ones in order to reach to a certain level of Quality of Service (QoS), at the lowest cost).

CR is still continuing to emerge as the "primary" technical solution for the scarce spectrum resources, in the future. "Cyclic recognition" of sensing, learning, and adaption in the CR enables it to "jump" between channels, according to time and spatial opportunities, per case. This way of rapid performance can, in turn, "affect" all related terms of accessibility, and influence both the information providers and the tracking spectrum users, therefore structuring a proper environment for the implementation and the development of the future "cognitive radio and network" (CRN) [2]. Software Defined Radio (SDR), on the other hand, is assessed as an "enabling technology" suitable for the introduction of a proper level of flexibility, allowing a device to become accustomed to its context.

Cognitive radio has been originally defined in 1999, by Joseph Mitola III, as "a radio that employs model-based reasoning to realize a specified level of competence in radio related domains" [3]. Following to this fundamental approach, several research and standardisation groups have further extended the essential CR concept,

occasionally by using the terms "adaptive", "cognitive", "learning" or "intelligent". In any case, a CR mainly consists of a cognitive engine (CE), which includes algorithms and toolboxes for radio environment sensing, machine-learning, and a configurable radio platform that could be an SDR. The CR concept was first described in detail by Mitola and Maguire in 1999 and it has been assessed as "transforming radio nodes from blind executors of predefined protocols to radio-domain-aware intelligent agents that search out ways to deliver the services that the user wants even if that user does not know how to obtain them" [4]. An ideal CR is able to have enough knowledge of its exterior surrounding environment (also including the radio environment) as well as of its potential user requirements. It has also the ability to influence further actions-processes and to make (selective) decisions about future design plans and/or to make negotiations to get access to the most appropriate part of the spectrum to achieve its operations within at the best power, modulation scheme and to handle such resource in real time, satisfying both the service offered and the users' demands [5].

According to the ITU official definition [6] a CR system is "a radio system employing technology that allows the system to obtain knowledge of its operational and geographical environment, to establish 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". In the same context, a software-define radio is "a radio transmitter and/or receiver employing a technology that allows the radio frequency (RF) operating parameters including, but not limited to, frequency range, modulation type, or output power to be set or altered by software, excluding changes to operating parameters which occur during the normal pre-installed and predetermined operation of a radio according to a system specification or standard".

In the following subsections we examine, per case, some among the core features of CR systems, that is: (i) The realization of a more efficient spectrum use and management; (ii) Enhancement of user experience; (iii) optimization of networks.

8.1.1 Realization of a More Efficient Use of Spectrum

The CR technology is a key concept, assessed as an indispensable "component" of the Future Internet (FI) and, particularly, of the imminent and beyond 4th generation (4G). Cognitive Radio Systems are expected to extend the efficiency of the complete spectrum usage by offering new sharing opportunities and are also able to deliver more flexibility to applications, due to their capability to adapt their operations to external and/or internal factors [7]. The related novel sharing prospects can help access to (new) spectrum bands; on the other hand, the increased flexibility to applications can improve cost efficiency, support delivery of services and promote users' needs [8]. Moreover, CR systems have the capability to increase the efficiency/flexibility of the radio resource management within wirelessmobile operator networks due to their dynamic ability to modify their behaviour and adapt their operations to any underlying environment [9].

In existing radio communications networks, all available frequencies are not in use, full-time. Based upon this consideration, CR systems aim to "recognize" the non-used portions of spectrum and to "share" these parts without interfering with other existing users harmfully, so that to influence the quality of the related services. Besides, spectrum usage can considerably vary over time and/or over different locations, depending upon specific requirements of use; when this can be the case, CR can "intervene" to improve efficiency and flexibility of spectrum usage. To realize this target, in essence the CR systems execute the following three processes: (i) They both perform collection and storage of knowledge of the radio operational environment and of the related operational location(s); (ii) They assess all previously gathered data and proceed by making a decision upon the collected information; (iii) Based on that decision, they further proceed by realizing specific actions. These three distinct and identifiable steps typically compose the well-known "cognitive" cycle (or the so called "MDE" (Monitoring/Decision-Making/Execution) cycle), already identified in a great diversity of implementations in the international literature [10].

The MDE cycle makes available a remarkable methodology for the design approach as well as for the proper modeling of scenarios and/or related use-cases of practical market interest [11]. The use of cognitive management for spectrum utilization is expected to increase the capacity of the networks, to permit access to new spectrum bands as well as to raise the economic value of spectrum. The introduction of cognition-based features in networks can offer means to overcome structural limitations of current infrastructures – which render it difficult to cope with a wide variety of networked applications, business models, edge devices and infrastructures – so as to guarantee higher levels of scalability, mobility, flexibility, security, reliability and robustness [12]. Dynamic spectrum access may occur in different ways, that is: (i) Between a licensed primary system and a licensed-exempt secondary system, for example secondary spectrum access to digital TV or military spectrum within the same primary system, (ii) micro-macro sharing of licensed spectrum in 3GLTE (long term evolution) femtocells and, finally, (iii) among primary systems such as real-time leasing and trading of spectrum between two cellular operators.

8.1.2   Enhancement of Users' Experience

One of the major targets of a CR-based network comprises the improvement of users' experience(s) to the extent possible. This is particularly important as it can offer appropriate guarantee for the provision of high quality seamless services. This can be implemented by taking into account miscellaneous situations, as below:

• Cross-Operator Access: A user can perform diverse subscriptions to a variety of services-applications offered by different market operators. Furthermore, a user can display different (favourite) preferences such as short download time,

stable connection and low cost. By further analyzing user preferences and user environment, the CR systems can help a user to connect its terminal device/equipment to the wireless access network that best fits his/her declared preferences.

- User Network: A user may hold a range of (terminal) devices and related communication equipment in his/her premises that need to be connected to each other and/or to the wider Internet. To realize a fitting (inter-)connection, different types of equipment may necessitate dissimilar prerequisites, (i.e., high bandwidth, low power consumption, etc.). An effective choice of operation parameters and/or protocols, can lead to the most appropriate way of connection, with the intention of better serving any expected functionality.

- Access-Flexibility to the Future Internet: Future systems are planned to offer adequate guarantee for continuous provision of (efficient) network access. Future Internet access requires new and innovative ways to access the network, route data flows to their destination, and access information. In this complex and fast evolutionary scope, the amount of required information is steadily increasing, also resulting in needs for higher data rates and improved spectrum utilization.

- Connectivity of smart spaces: Apart from the requirement to offer a quick and reliable access to a central underlying network, the number of all corresponding interconnected devices can grow exponentially. In the networks of the future, miscellaneous devices should be able to interconnect wirelessly, thus implying the necessity for fast device discovery and for an "agile" use of the spectrum to ease ad-hoc interactions between devices and long life-times of networks, under uncertain connectivity situation.

8.1.3 Optimization of Networks

A core challenge, for a network operator, is to be able to respond to most – if not all – identified needs of his network users in a timely manner, particularly by satisfying all related requirements in terms of capacity (i.e., throughputs) and quality of service (QoS). CR systems can collect informative data from the radio surrounding eco-system and, then to analyse/assess the corresponding radio-operational environment, so that to make the network itself capable of suitably reacting to a certain state, that is by "optimizing" the selection of radio access technologies and of associated radio resources [13]. This can occur to support the optimization of a wireless/mobile operator network, to effectively deal with the following issues:

- Load balancing: Load balancing mechanisms can be a critical part of the intended full optimization activity, especially in cases of traffic variations in space/time or nature of required service/application. Uplink traffic demand information from the user equipment (UE) and/or related sensors can "trigger" the obligatory reconfiguration activity of the radio elements both in the network and user (multimode) terminals.

- Spectrum refarming: A case of interest is the one of spectrum refarming, specifically within the framework of technological evolution and/or periodical appearance of new families of standards. This necessitates their progressive introduction/coexistence in the legacy "bands" rather than performing an "easy and rapid switchover" which is not suitable because of the large quantity of legacy equipment and the corresponding investments [14]. CR system can facilitate a "smooth" refarming transition period, by assessing and evaluating traffic constraints and user requirements.
- Radio Resource Management (RRM) optimization: If we consider a base station (or a cell) that has been located in a certain site and covers a specific area, it can be easily observed that the traffic of different services upon a specified RAT may change from one sub-area to the other according to the day period and depending on different factors (e.g., number of active users, kind of requested services, related bit-rates for downloading, etc.). Furthermore, it could happen that some cells may be congested (high blocking percentages) in some specific area(s) (usually these portions are called as "hot-spots") in which the traffic is more consistent, while some surrounding cells are "less loaded" or may be characterized by low blocking percentages. Besides, in case of the deployment of two or more RATs simultaneously in the same area, the offered traffic of different services on each deployed RAT may also be differently distributed in time and space, with respect to the ones of the other deployed RAT(s). In such complex contexts, in the longer term, CR systems can give the network operators the means for managing in an efficient way the radio resources existing within their own licensed frequency bands.
- Cognition Enabler: Allowing for an heterogeneous (or multi-RAT) context in which Dynamic Spectrum Allocation (DSA) schemes could emerge, the corresponding mobile/wireless terminal has to start a communication in a spectrum context which is "mostly unknown" due to such dynamic allocation mechanisms, without requiring an excessive complexity. In this case, it is essential to have well-organized mechanisms to provide the adequate and exact information to the terminals, for the latter to be able to start a communication session, appropriately. One of the most vigorous areas of present research in CR is focused upon DSA, aiming to detect and to access "spectrum holes" of "white spaces" [15]. The target is to accomplish device-centric interference control and dynamic reuse of the radio spectrum, based on the frequency agility and the intelligence which is created by the CR technology.

  One among the present critical challenges for the research and the industrial community is to "bridge the gap" between individual research results and the large-scale deployment of CRN. This implies the need for steady efforts to allocate for technological solutions capable of taking advantage(s) of the enormous potential and commercial promises of CRNs [16]. In a longer timeframe, it is expected that DSA will exceed the opportunistic spectrum access model and new technologies – or policies – will be established for CRNs, to access a portfolio of different spectrum types [17].

Typical examples comprise licensed spectrum, unlicensed spectrum, and leased spectrum. The majority of future devices are expected to be able to dynamically change the operating spectrum within the particular spectrum portfolio, as well as to realize/execute this action on a "just-in-time" basis. Of course, the resources of the "spectrum pool" may depend on context, location, and technology. Parameters like price, QoS and energy-saving may be used for the selection of the particular spectrum, to serve specific purpose(s).

One other significant future research direction is associated with the advance of national and/or international interactive spectrum use "maps", able to give information about the spectrum availability in different places of the world. These maps will further allow for the evaluation and tuning of other key technologies (like, e.g., wide-band antennas, spectrum sensors).

- Decentralized RRM: The intension, here, is to "lighten" the complexity of the system by "extracting" a part of the radio resource management functionality into more decentralized solutions. To realize this, essential context information (e.g. network radio capabilities, network/mobile measurements, geo-location information, etc.), network policies and other kind of information (e.g. operative software to reconfigure the terminal or advertising or roadcar traffic information) have to be adequately provided to all interested devices/equipment.

8.1.4 Technical Requirements on Current CR Systems, as Identified by Actual Standardization Works

In this part we discuss, in short, some key-requirements for CR systems as these have been identified and examined by actual ETSI works [18, 19]. Technical solutions to meet these requirements are indispensable to provide cognitive radio services for wireless/mobile users. These are listed as follows:

- Scalability and Insensitivity to Network Topology Changes: The cognitive radio networks scale well with the number of usersnodes in the networks. The cognitive radio network reacts "gracefully" and is robust to changes in network topology. Links may disappear because of agile spectrum conditions, and nodes may leave the network, for instance because of user mobility or powering off devices. Connectivity between the network nodes should be maintained in a robust manner, and advanced protocols are required that reconnect nodes via different frequency bands.
- Power Efficiency: The power consumption of the CR system design allows long standby, monitoring and active times. As a whole, this prerequisite is expected to be fulfilled by each network node, when taking into account also the new functionalities like relaying, collaboration and cooperative spectrum sensing.
- Network and Service Discovery: The CR system provides the first connection to the desired service in a short time, also being acceptable to a customer. A user can "join" a network within a reasonable time. The CRS protocols are designed

in such a way that the user can be informed in reasonable intervals of delays and progress of a connection/service request, if a certain deadline defined by network policies is not met.

- Robust Control Plane: The control planes – both within all cognitive radio networks as well as between them – should be robust and able to continue to provide connectivity, in frequency agile environments.
- Reconfigurability of the Radio Equipments: The radio equipments in the cognitive radio network nodes should be capable of adjusting to different radio frequency environments. This kind of frequency agility means that the transmission parameters and resource allocation can be easily adjusted to the needs of the operator and/or the user, or the interference environment in a particular band.
- Context, policies and information provisioning support: For the purpose of supporting cognitive radio network nodes in their selection of radio technology and frequency band as well as radio link configuration, context provision needs to offer the radio context information such as available frequencies and radio technology selection constraints (policies).

8.1.5 CRNs for Further Network Deployment

The research activities in the domain of CRNs are established by considering both the user and the network operator demands. For example, some of the critical expectations of users are relevant to a diversity of factors, including but not limited to: higher bandwidth demand; user friendliness and personalization; user-centric communication; seamless integration of already existing and new terminals, networks and services; security facilities; cooperative communication by customer diversity and; customer accumulation to improve the performance as well as attractive billing systems. On the other hand, operators look forward to develop solutions for example on better handling the network management complexity together with network security and scalability provisions, fault tolerance, fast integration of new technologies as well as attractive – and easy to be offered to the market-business models [20].

The complexity and diversity of such demands can influence, in turn, research activities in CRNs, thus imposing several priorities [21] for further evolution in thematic sectors, such as: (i) Improvement of existing techniques and algorithms to optimize the performance [22]; (ii) Development of new – and more enhanced – protocols; (iii) Scientific approach, which demands for the development of simulation and/or analytical theoretic models to improve existing techniques, and; (iv) Breaking the protocol layering and bringing in cognition facilities in networking and wireless communications.

A significant issue in the present research effort on CRNs is that a "n-dimensional" space could be developed and also be used for communication and, consequently, it should be able to exploit "under-utilized" spectrum portions. Examples of potential "dimensions" with realistic significance and meaning for

the market are: frequency, power, geographical space, code, angle and time. The consideration of some among these in the process for identifying spectrum holes, can guide to modern management, routing and architectural for the proper CRNs' management, implemented at the application layer [23]. Operations like spectrum sensing, addressing, decision making, routing, handover and prediction can be done in an "n-dimensional" space and be used for knowledge representation and management of CRNs.

In the existing wireless networks, people select and use a certain wireless network by subscribing to a particular wireless network service provider. This approach has worked well so far, but it is expected to "face" major problems in the future, given that the wireless network service providers focus on achieving the best performance in their own network only. Providers generally ignore the effects on other (adjacent) networks, usually sharing the same operational area(s). A straight outcome of this concern is that any independent network design and operation effort increase the inter-cell interference among homogeneous and heterogeneous networks and does not provide for an optimal use of the radio resources, for all the underlying networks. In addition, the locations of the base stations need to be selected very carefully, with the help of cell planning and should be configured so that to "maximize" the corresponding cell coverage. However, deployment is enormously complex and the location acquisition for base stations is very difficult, especially in urban areas (e.g., in case of high towers).

Therefore, a more flexible deployment is needed for the service providers to easily and cost-efficiently plan the cell. To achieve this aim, cooperative heterogeneous networks could be examined as a potential "candidate" solution to make flexible and low cost deployment by using macro-, pico-, femto- and relay stations in combination with high capacity. The upcoming and beyond 4G wireless systems are thus expected to improve the performance through internetworking and cooperation, especially if they also take into consideration options for an "optimal" radio resource use – for example, by investigating and establishing CRNs with cooperative heterogeneous networking [24]. Basically, CRs enable mobile stations to reconfigure the transmission parameters so that they can share unused licensed spectrum. The CRN intelligently changes the network parameters based on radio environments so that it can achieve more efficient spectrum use.

8.1.6   Conclusion

Cognitive Radio Networks (CRNs) are emerging as a viable solution to solve spectrum shortage problems, with the view of the evolutionary progress towards realizing the network of the future. In particular, by configuring different transmission parameters (e.g., frequency band, waveform, transmit power) on-the-fly, wireless and mobile networks resources can be shared with licensed users in an opportunistic manner, to offer the best possible result.

In this work, we have provided a general survey of several CRNs, as identified by actual European standardization efforts, including the critical capabilities andor

features characterizing them. To this aim, an overall and harmonized technical concept for future CR systems has also been outlined. There are several factors driving the future evolution of radio technologies and network architectures towards more flexible and reconfigurable CR systems, as the latter are expected to increase the efficiency of the overall spectrum usage by offering new sharing opportunities and also to provide more flexibility to offered services-applications-facilities.

The new sharing possibilities can facilitate access to new spectrum bands and the increased flexibility to applications can improve the cost efficiency and capabilities to deliver various services as well as better take into account the users' needs. CR systems are also expected to increase the efficiency and flexibility of the radio resource management within (wireless/mobile) operator networks (as a result of their ability to adapt their operations to external and internal factors).

8.2 Indoor Positioning and Horizontal Handover Employing RSSI Fingerprinting in OTE Labs

Nowadays it is very essential for an operator to be able to locate a user in indoor environments and offer him the same Location-Based Services (LBS) which are offered to him in outdoor environments. Therefore, LBS services can be further extended in indoor environments if reliable and reasonably accurate three-dimensional positional information of a mobile device can be determined seamlessly in both indoor – outdoor environments and in cases of movements between the two environments. There are situations that the network, being either wireless or fixed needs to be improved in case that there is either external interference or high throughput use in a link or in a communication route and the traffic needs to be handed out to another node or to another network.

In addition, location awareness is a very significant characteristic of the cognitive networks in terms of energy efficiency and throughput improvement. In indoor environments it is crucial to be able to estimate the location of a user with good accuracy. This report presents the performance and the accuracy achieved in estimating the user's position in indoor environments by exploiting the fingerprinting method. This can later be used for purposes of network improvement or better service delivery.

In this direction, there are many efforts in providing algorithms and techniques for accurately estimating the geo-location of a user in an indoor environment. Geo-location for cellular networks represents a great interest and a major importance for telecom operators and for users who are eager to use context aware services. It can be used in many applications such as emergency, security, tracking, monitoring, intelligent transportation systems and energy efficiency in base stations and mobile terminals [25].

Location based services (LBS) are widely recognized as value added services and due to the diversity of user requirements research efforts are needed to improve the location determination capability and its accuracy and reliability. The most popular commercial solution reliable on the market to get accurate location information of

the cellular phone is the Global Positioning System (GPS), where time of arrival (TOA) [26], or time difference of arrivals (TDOA) [27] measurements are calculated to provide the localization service.

However, not always the GPS is the most suitable solution for localization. Its main limitation is seen in severely handicapped environments, such as outdoor urban canyons and indoor environments, which actually represent the greatest interest of service providers. In these conditions it is difficult, if not impossible, to obtain any sort of location information.

This is due to the infeasibility of having a clear view of at least four satellites, or due to signal blocking and multipath conditions. When localization is done in cellular networks, the accuracy is highly dependent on the channel and line of sight (LOS) conditions. In some cases, due to the channel conditions poor results are obtained.

The localization accuracy in cellular networks can be improved by exploiting both long-range and short-range technologies. This can be done by cooperation between users using WiFi communication in short range, exchanging their location information which is embedded in the received power. Nevertheless, cooperation must be assured by means of incentive schemes and security mechanisms that will deter users from defecting, e.g., hiding or sending erroneous information. In this case, by exploiting the WiFi technology for positioning estimation, the so-called fingerprinting method based on WiFi signal strength observations is usually employed. This method has been used also by other investigators who report results on the accuracy of the method [28–31].

8.2.1 Positioning Technique

In this paper we apply the fingerprinting method in a WiFi 802.11 environment in order to determine accurately the position of a user and use coverage data from a ray tracing tool in order to compare the accuracy of the method. This method can be used to accommodate cognitive networks in indoor or outdoor environments.

The fingerprinting method has been proved as an effective method for determining the position of a mobile user in an indoor environment. Several researchers have found fingerprinting as an effective method for user's location estimation in indoor environments by utilizing WiFi's radio characteristics such as the RSSI and bit rate of transmitted traffic and the positioning phase where the position of a terminal is estimated by exploiting the saved data in the database.

During the learning phase, a database is constructed that consists of a radio and a performance map where each specific location point corresponds to a value of RSSI and bit rate (BR) data at specific points which are used for estimating the best location of the user. Therefore for each location point, there is a set of reference values regarding the location of the test user, the radio map data RSSI and the performance data of bit rate. For the second part of the procedure where the positioning or location estimation of the mobile user is determined, then measured
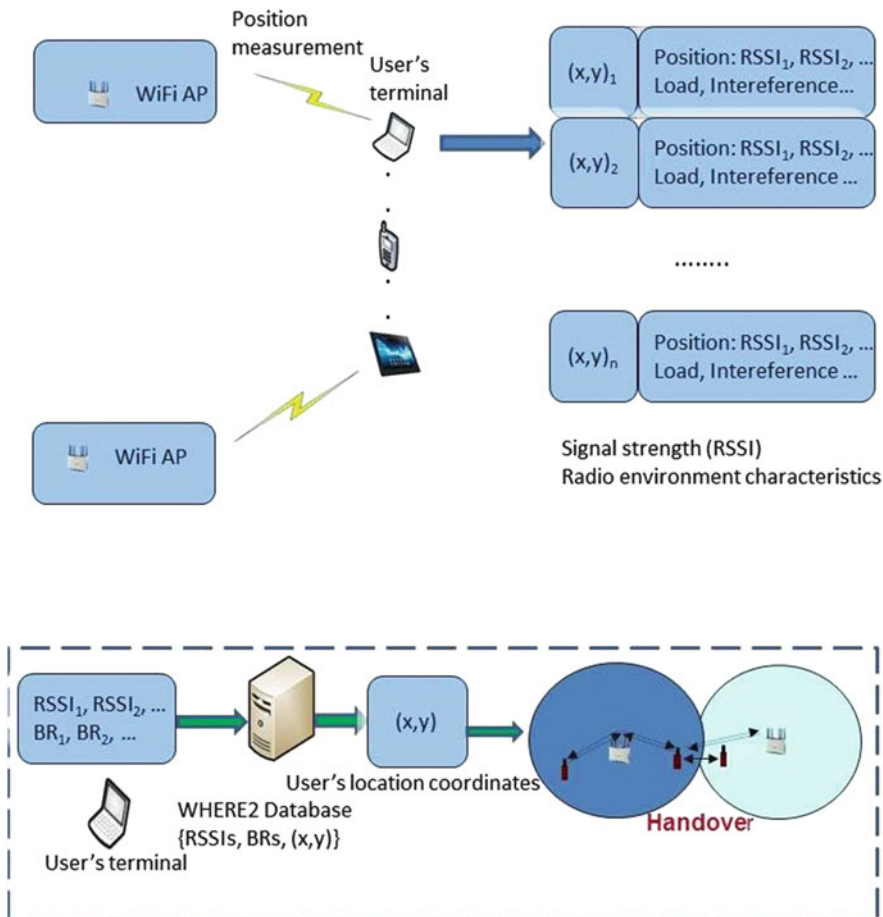
Fig. 8.1 Phases of the fingerprinting method

data of the SNR and BR at specific points are performed and the position of the user's terminal is estimated by matching the above measured value with the ones reserved in the fingerprint database.

The fingerprinting approach has been proved to be an effective method for WiFi positioning although there are still a lot of problems such as interference from other nodes or multipath fading from different obstacles in an indoor environment. The fingerprinting method works in two phases, where during the first phase the WiFi nodes are placed in different positions and signal strengths are obtained by moving a user's terminal around the area. The average signal strength (SS) of each WiFi AP measured at each point and the resulting data are used in order to create the fingerprint database.

The procedure of the fingerprinting method is depicted in Fig. 8.1. Since the variation of the SS measured at each point is large due to the above reasons, a
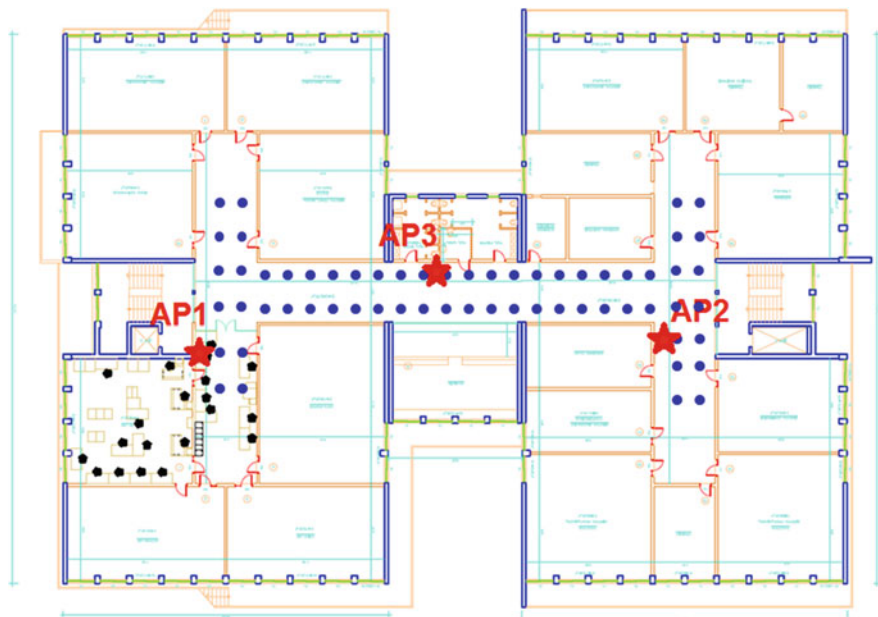
Fig. 8.2 Indoor environment of OTE Labs (first floor)

probabilistic approach has also been developed in order to approximate the SS in every point in a layout of a building as the one shown in Fig. 8.2. The above figure shows the floor layout for which the fingerprinting method was used by exploiting three WiFi APs and a mobile user with a laptop as a terminal with WiFi 802.11g/n. Since the variation of the RSSI and BR measured at each point is in some cases considerable, which is affected by the environmental noise or by the movement of people who cause multi-path effects, a probabilistic approach has also been developed in order to approximate the RSSI and BR in some typical points of the floor layout.

In such a way the localization means of GPS can be used for outdoor environments where WiFi is used for indoor ones.

8.2.2 Experimental Setup

In this section, we present a use-case of indoor positioning based on RSSI fingerprinting methods in the indoor office environment of the OTE Labs building (first floor), which is depicted in Fig. 8.2 as mentioned before. In this environment, we installed three commercial Cisco Access Points (APs) in order to offer WiFi coverage (2.4 GHz) to the corridors of the floor. The common SSID of all three APs was "WHERE2", while each of them was set to use a different frequency channel, i.e. $AP1$ used channel 1, $AP2$ channel 11, and $AP3$ channel 6.

8.2.3   Indoor Positioning Application

In order to implement the indoor positioning, we developed an application for the Android OS, which interacts with a MySQL database, which has been deployed to store RSSI measurements, position and other related information. Concerning the implementation details, we used 2 ZTE Light Tab 2 terminals the Eclipse IDE with the Android SDK 4.2 compiler, while we created a MySQL database using XAMPP 1.8.1 to store the measured RSSI values.

The Android application, which is titled "Where2", employs the following features:

- Finger: scans the transmitted RSSIs from the 3 Where2 APs including related channels and frequencies, MAC addresses, etc., and records them in the MySQL database,
- Change: scans the transmitted RSSIs from the 3 Where2 APs and enforces the connection of the terminal to the AP with the highest RSSI value (i.e. best signal) (horizontal handover)
- Status: presents in the screen information about the current connection,
- Find: runs the RSSI fingerprinting algorithm, identifies the position of the mobile terminal and displays it onto the map (see Fig. 8.3).

Figure 8.3 depicts the home screen of the Where2 application as it appears in a ZTE Light Tab 2 with Android 2.3.5.

8.2.4   Status and Finger Operations

An instantiation of the Where2 application for indoor positioning can be seen in Fig. 8.4, while the successful insertion of the RSSI information is demonstrated in Fig. 8.5.

Specifically, after the initialization of the application, the mobile user taps the Status button to derive information about his current connection. Then, the mobile user taps twice the "Finger" button to identify RSSIs from neighbouring APs and to record the received RSSI values in the MySQL database, which is named `rssi_positioning2`. In Fig. 8.5 entries of the database are illustrated, where the MAC address of the terminal device and the three RSSI values are included (e.g. fields highlighted by two ellipsoids).

8.2.5   Find Operation

In order to accommodate the successful indoor positioning of the mobile terminal, we initially deployed a database called rssi_finger, where a multitude of RSSI 3-tuples along with related (x,y,z)-coordinates were stored. These values were
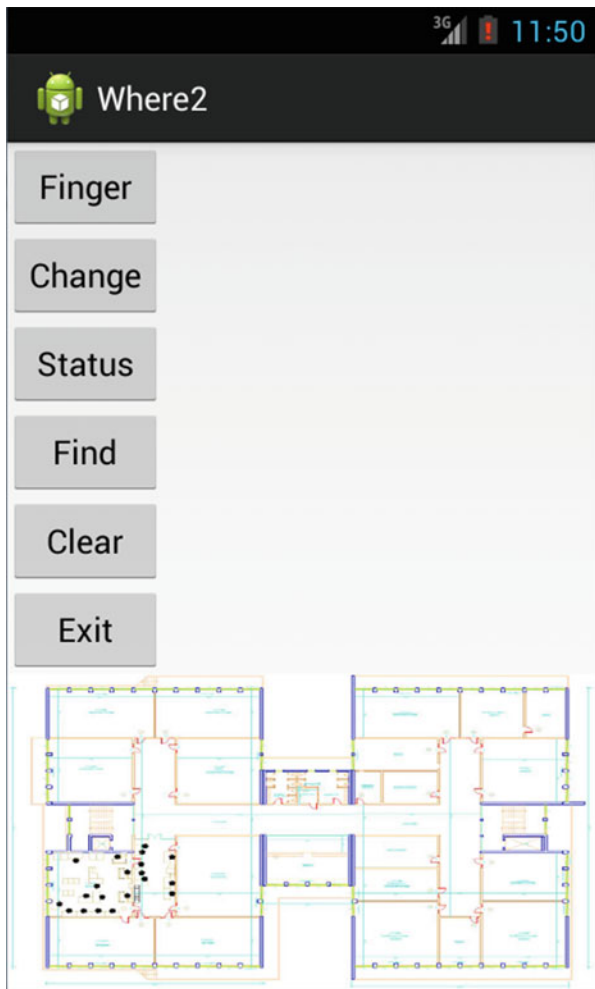
Fig. 8.3 Home screen of the Where2 application for indoor positioning

derived by measurements taken throughout the entire WiFi coverage area, i.e. pointed out with dots in Fig. 8.2.

Then, the RSSI fingerprinting algorithm for the indoor positioning performed the following steps:

- Scan the current RSSI values and store them in the rssi_positioning2 database,
- Calculate the mean least square error between the current RSSI values and each of the recorded RSSI 3-tuples in the rssi_finger database (also called training database), and
- Return as current_position the (x,y,z)-coordinates of the recorded RSSI 3-tuples with the minimum "distance" to the currently measured RSSI values.

Fig. 8.4 Instantiation of the Where2 application for indoor positioning

Fig. 8.5 Instantiation of rssi_positioning2 database successful insertion of RSSI fingerprinting information

```
while($row = mysql_fetch_array($result)){
    $aData[]=$row;//array with all radiomap points in database
}
$sData = array("$ap1","$ap2","$ap3");// the 3 RSSI from user device
$count = mysql_num_rows($result);//total radiomap points in database
    for($i=0; $i<$count; $i++){
        $temp3=0;
        for ($k=1; $k<4; $k++){
            $temp1 = $aData[$i][$k] - $sData[$k-1];
            $temp5 = pow($temp1, 2);
            $temp3=$temp3+$temp5;
            $tempar[$i] = $temp3;// array to find radiomap point id from database
        }
    }
$temp=$tempar[0];
    for($i=0;$i<$count;$i++){
        if($tempar[$i]<$temp){
            $temp=$tempar[$i];
            $pos=$i+1;//the nearest position id
        }
    }
```

Fig. 8.6 PHP source code for identifying the "closest" record w.r.t. RSSI

```
/*the below query is to send to user device the coords
 in order to display a dot on map image*/
$query2="SELECT `X`,`Y` FROM rssi_finger WHERE `ID`=$pos";
$result2=mysql_query($query2);
while($row = mysql_fetch_array($result2)){
    $dData[]=$row;
}
$temp1=$dData[0][0];
$temp2=$dData[0][1];
echo "$temp1 s $temp2";//return coords
mysql_close($mysql_link);
```

Fig. 8.7 PHP source code for returning the estimated indoor position to the mobile terminal

The "distance" metric employed to calculate the difference of two RSSI 3-tuples is:

$$\min_{i=1,\ldots,m} \left| \sqrt{\sum_{i=1}^{m}(SS_{RM}(i,j) - SS_{MEAS}(i))^2} \right|, j = 1,\ldots,n, \qquad (8.1)$$

where $SS_{RM}(i,j)$ is the SignalStrength value of the signal transmitted from access point i at radiomap point j, and is the measured SignalStrength of the signal transmitted from access point i. The radiomap point j having the minimum norm is considered to be the most probable location.

Finally, Figs. 8.6 and 8.7 present the PHP code running in the XAMPP server side in order to calculate the indoor position and return it to the mobile terminal, respectively.
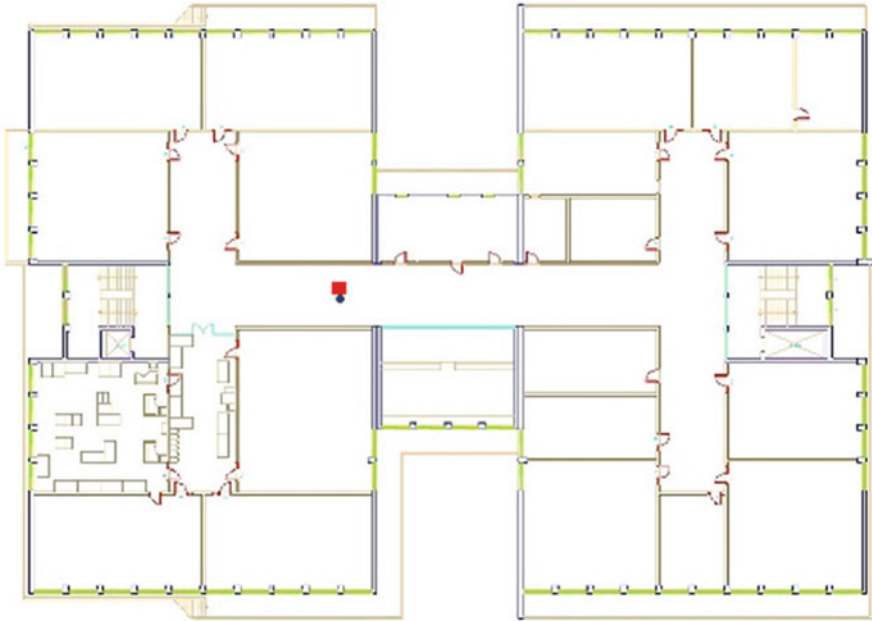
Fig. 8.8 The small cycle (dot) represents the actual position and the larger rectangle the estimated one

Finally, the outcome of a Find operation is then presented to the screen of the mobile terminal.

8.2.6 Accuracy of Indoor Positioning

An example of indoor positioning by means of the "Where2" application taking place in the indoor environment of OTE Labs is depicted in Fig. 8.8, where the small cycle (dot) represents the actual position and the larger rectangle the estimated one.

In general, the error of the RSSI fingerprinting method is depending on the circumstances, e.g. glass and metallic surfaces, thick walls, etc. Figure 8.9 shows the accuracy on the predicted position data of a user in an indoor environment by using the fingerprinting method. It is observed that the accuracy varies between 2.0 and 3.3 m, which is higher compared to the expected values, i.e., from 0.5 to 2.5 m. This is due to the inability of the method to overcome problems such as multi-path in NLOS environments. More advanced techniques such as diversity, directionality with beamforming antennas, positioning cooperation with other technologies are needed in order to obtain more accurate positioning measurements.
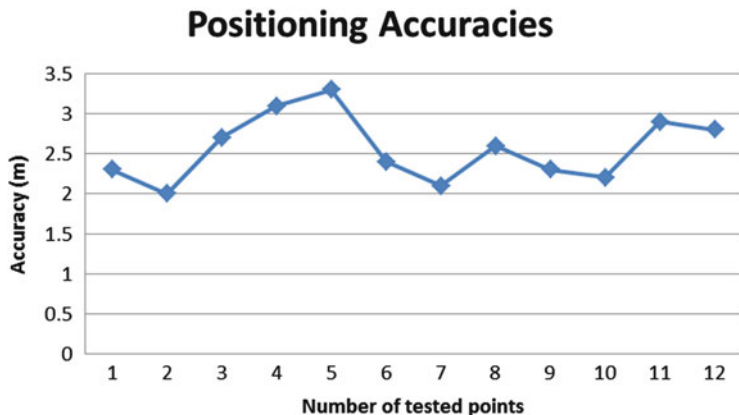
Fig. 8.9 Accuracy of the predicted position of a user using the "Where2" fingerprinting application

8.2.7 Conclusion

In this paper, we present a location estimation application based on RSSI fingerprinting, which is used for estimating the geolocation of a user in an indoor environment. The so-called "Where2" application is modular, and can be easily extended to employ alternative metrics for the calculation of the indoor position.

Moreover, the estimated indoor position can be employed to perform horizontal handover, e.g. consider also geographical distance, when deciding which AP to connect to. Additionally, the derived positioning data can be used for improving end-user's Quality-of-Experience (QoE), achieving energy efficiency and supporting personalized and location-specific advertisement.

Further studies and simulation results will be provided as a future work. Specifically, the contribution of the fingerprinting application will be examined in an indoor environment in collaboration with cognitive and load balancing techniques. The performance of the network will be examined under critical key performance indicators, such as network resource utilization and QoS/QoE indicators for specific services, e.g. voice, video, etc. offered to the end-user.

8.3 Cognitive Cooperative Positioning

The objective of cognitive radio is to define means on how to share the spectrum resources between different systems. A common example is that a primary user who owns the right to use the spectrum is sharing this with secondary users. Spectrum itself is a scarce resource, but the need to use itself is roughly limited
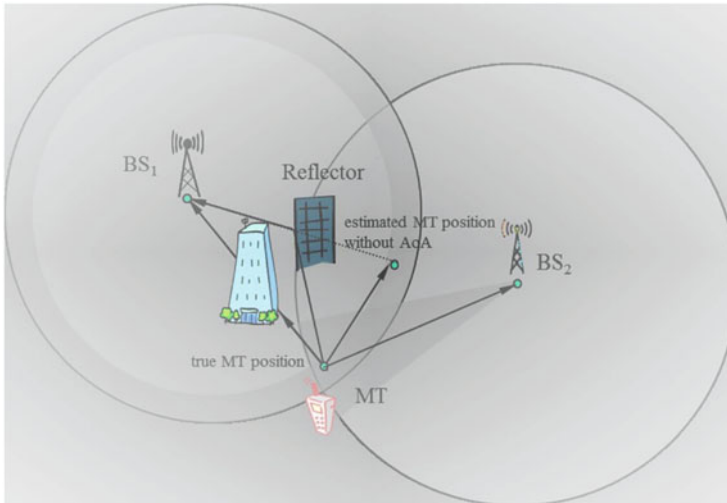
Fig. 8.10 Combination of ranging and angular information determines the true position of the mobile terminal (*MT*) in the cellular network

to the dedicated space between transmitter and receiver. Two different approaches have been developed for secondary users to access the wireless channel without severe consequences for the primary user. The secondary user senses the wireless channel to learn if the radio devices harm or interfere with a primary user. The alternative approach demands that the secondary user determines its geo-location to interact with a database that announces which spectrum is free to use under dedicated conditions. With this the cooperation between mobile terminals helps to support cognitive radio networks by improving the geo-location performance especially indoors, and to exchange and update the applied radio maps.

8.3.1 Geo-Localization

There are several ways to localize a mobile terminal in wireless networks through radio signals. In the following subsection we briefly introduce the basics of localization. Trilateration determines the position by measuring the distances between several radio nodes. This can be achieved either through the estimation of the transmission time between synchronized transmitters and receivers, or by mapping the signal loss between transmitter and receiver according to a path loss model to a dedicated distance. By solving the geometric problem, e.g., through the intersecting circles that each represent the distances between the transmitter and receiver. Triangulation exploits angular information, given from sectorized antennas or antenna arrays that are able determine the angle of arrival, to determine the position of a radio node. Figure 8.10 shows a combination of both, trilateration and

triangulation, to reduce any erroneous impact on determining the location of the mobile terminal. Most humans use their mobile terminals indoors. Therefore, the coverage of multiple base stations indoors is mandatory to successfully position the terminal through trilateration. The external or internal base stations act as anchors, but because of the rich multipath conditions between terminal and anchor accurate positioning is challenging. Additionally, we propose to enhance the connectivity by using peer-to-peer (P2P) links between mobile terminals. Mobile terminals are often close by other mobile terminals and with this we improve the coverage of such terminals even deep indoors by building an internal network that allows us to position the terminals indoors by the P2P links. The P2P links could use an additional air interface something which is commonly onboard on today's smartphones. The frequency bands of the operator could be reused internally to range or even communicate between the terminals.

8.3.2   WHERE2 Project

In the WHERE2 project [32] the focus was on indoor positioning mainly by mobile terminals with different air interfaces. Indoors mobile users usually enjoy the potential in terms of mobile-to-mobile cooperation and heterogeneous access to various radio access technologies (e.g. with respect of multiple access points). The positioning research in WHERE2 addressed multiple avenues, such as indoor channel modeling for positioning [33], evaluating inertial sensors with advanced fingerprinting methods [34], and learning and characterizing the environment based on the received radio waves jointly with advanced mobility models to localize the terminal as well [35].

In the following subsection we focus on two key topics that are outlined in more detail in [36]. First, message passing algorithms define how and which information should be exchanged (e.g. among the mobile terminals) in a distributed cooperative localization system. Different algorithms show specific benefits in different network topologies and densities. Message passing algorithms cause higher latency due to the slow convergence as several packets need to be exchanged compared to conventional algorithms, such as centralized trilateration. However, the additional information by the local neighbors proves the additional effort is worth being paid. The drawback of the latency due to the convergence is the lower adaptability of message passing algorithms in mobile scenarios. An accuracy of below 1 m indoors can be achieved by message passing algorithms, and which outperforms the accuracy of the least-squares centralized solution in some deployment and measurements scenarios. Thus, the cooperative localization performance is much better than the expected GPS performance outdoors [37].

Second, the rich amount of potential links between the devices causes problems due to interference. Especially if secondary users can access the wireless channel freely and without any controlled access scheme. Therefore, a synergetic mechanism between communication means and localization requirements
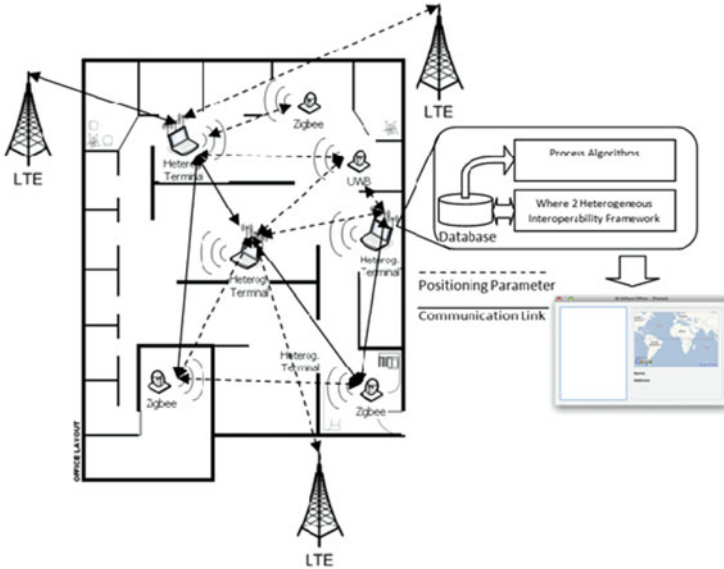
Fig. 8.11 The setup of a heterogenous cooperative positioning system. Different radio nodes with one or multiple air interfaces interact and share their estimated ranges and positions through a centralized database. This allows the terminal in the middle of the map to localize itself by connecting to only a single cellular base station

((un-)coordinated or (de-)centralized) could avoid a diminishing performance. Figure 8.11 shows a heterogeneous cooperative positioning setup.

Therefore, a link selection algorithm determines out of all available links the profitable links to reduce complex calculations in the mobile node to the lowest level according to the accuracy requirements. Different criteria, such as the individual ranging quality, the interference impact on other radio links, the geometrical constellation, and the theoretical accuracy predictions are evaluated to optimize the link selection process. Finally, depending on the prior knowledge the link selection is performed either at the transmitter (by avoiding any interference in the network) or at the receiver (cancelling). In cooperative positioning by message passing algorithm we consider a network graph that is not complete and as a consequence, there is maybe not enough information to estimate the different locations without ambiguities. We improve the detection of the uniquely solvable node, and we derive relationships between the concepts of rigidity and identifiability which allows us to exploit identifiability properties by using graph rigidity tools.

Finally, we tested some of our theoretically developed algorithms by using two different sets of data. First by the ray-tracing data of the Siradel environment that was generated by ray-tracing software. Furthermore, we used collected real measurement data of two radio access technologies that were developed within the project. The data was applied to evaluate and compare different message passing algorithms – which is the first comparison of the different algorithms itself – and
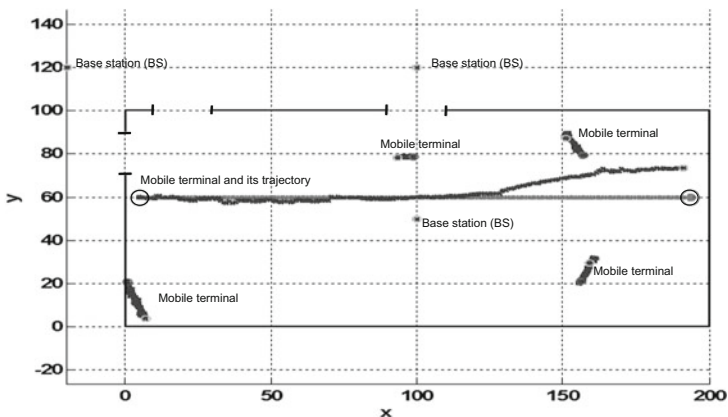
Fig. 8.12 An indoor scenario with five mobile terminals and three base stations (one indoors). The indoor area has three openings or windows that offer line-of-sight conditions between mobile terminal and base stations if applicable. One mobile terminal moves from left to right and is connected with multiple other terminals. The estimated trajectory suffers from the ambiguity of the two mobile terminals which are in an orthogonal direction to them. Further, the out base stations move from a line-of-sight connection to a non-line-of-sight connection
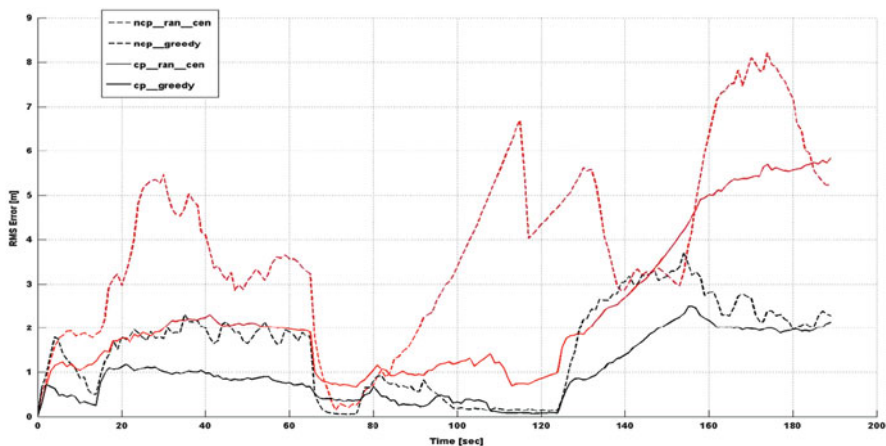
Fig. 8.13 Performance comparison between cooperative and non-cooperative positioning. The spectrum resource is allocated according to different criteria, such as the geometric constellation and the path loss, to the different mobile users

different link selection techniques. Figure 8.12 shows the scenario of five mobile terminals, with one moving and four cooperative terminals. Three base stations or anchors support the determination of the location of the mobile terminals. Figure 8.13 shows the performance results for non-cooperative and cooperative algorithms that allocate individually spectrum to the different mobile terminals depending on the algorithm. A random algorithm allocates spectrum randomly to

the different users. A greedy algorithm considers in the spectrum allocation next to the path loss, also the geometric constellation of the different radio nodes including the base stations to each other.

Especially, the real measurement data [38] showed the intrinsic problem of using real devices, namely the very limited data samples caused problems in deriving statistical models, and the severe outliers complicated the modeling further. Nevertheless, jointly with the environment we applied successfully hybrid data fusion exploiting the LoS and NLoS cases and could show the benefit of cooperative and decentralized over non-cooperative positioning. We see significant benefits from contextually selective hybrid data fusion (HDF) in NLOS vs. LOS.

8.3.3   Multi-user Positioning

We consider the (perfectly known) positioning CRLB to allocate frequency resources to enable simultaneous multiuser positioning [39]. The key idea is that the total spectrum shall be divided into portions so that multiple devices can access a channel simultaneously. A simultaneous access of the channel to determine the location reduces the latency compared to classical location technologies that rely on the protocol stack of a communication system. From the TOA CRLB we can see that different subspectra (or in OFDM subcarrier) have an unequal impact for the ranging performance of each link. Jointly with the geometric constellation of the nodes to each other we applied different allocation strategies for different subspectra to be used by different users at the same time. The different methodologies to allocate the subspectra investigated are game theory (Nash equilibrium), the greedy algorithm and random allocation. Further, we consider a centralized approach that follows the constraint to avoid any interference. Alternatively we follow a distributed setup with either partial or no knowledge about neighboring mobile terminals. Therefore, we allow interference of the ranging signals. The centralized approaches avoid any interference by neighboring nodes simply by avoidance of the same spectrum. This is beneficial in dense networks with low distances. However, if the impact of pathloss on the wireless signal is severe enough manifold assignment of the same spectra and allowing low interference is more beneficial than using less spectrum or the less significant spectrum. The distributed approach suffers from two drawbacks. First the interference degree is unknown and is either considered as a fixed constraint or is not considered at all. Allowing interference in a dense network could still be beneficial as it allows to use neighboring agents to allocate more significant spectrum as well. A revised ranging and positioning CRLB is derived to consider the signal to interference and noise ratio for allocating the subspectra. Second the knowledge of uncertainty of the estimated ranges from the neighboring nodes is considered to reduce the impact of error propagation [40].

Jointly with the subcarrier allocation we developed an enhanced particle filter for tracking agents by integrating a multi-hypothesis mobility model based on the
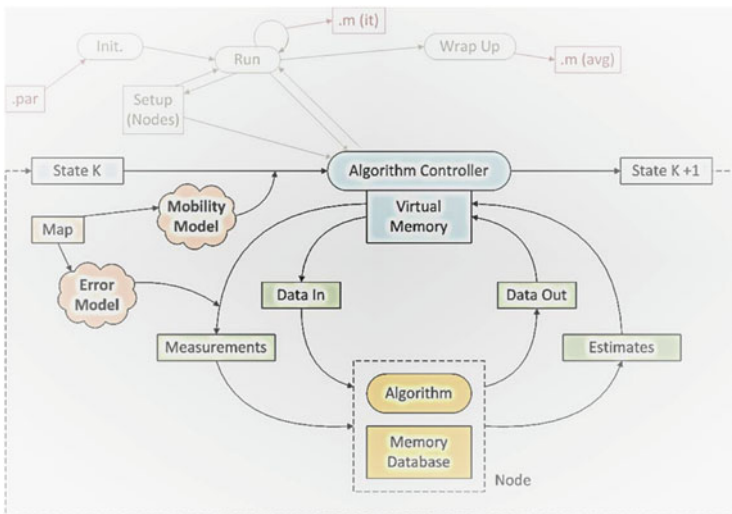
Fig. 8.14 The general structure of the cooperative and distributed positioning simulator

Levy flight distribution [39, 41]. With this we are able to cover typical nomadic movements of the mobile agents such as stays or walks. The advantage of the multi-hypothesis tracking algorithm is the lower risk of losing the track together with the advantage of reducing the ranging variance – both results in a lower root mean square positioning error. Figure 8.14 illustrates the structure of the cooperative and distributed simulator. The simulator is written in Java and operates in a successive mode. In each state different information is considered. It starts with the algorithm controller which requests input about the estimated location of the neighboring radio nodes or the ranging estimates of the neighboring radio nodes. Further, a general error model considers ranging errors between the radio nodes depending on line-of-sight or non-line-of-sight conditions, mapping information and mobility behaviour. All this is considered in the different potential positioning algorithms that the simulator can select from. The memory database allows to select which information is locally available. Therefore, we can differ in our simulations between fully distributed and centralized algorithms. The generated output is prepared for visualization in Matlab.

In Fig. 8.15 we compare the performance of cooperative positioning vs. non-cooperative positioning. The bigger dots represent anchors and the smaller dots represent mobile terminals. The mobile terminals try to determine their position by using the transmitted signals from the anchors. The light gray (outer) ellipsoids represent the expected lower bounded positioning performance in the different dimensions. The bold (inner) ellipsoids represent the performance by incorporating the signals from the other mobile terminals (smaller dots). The improvement by the cooperative setup is significant comparing to the non-cooperative.
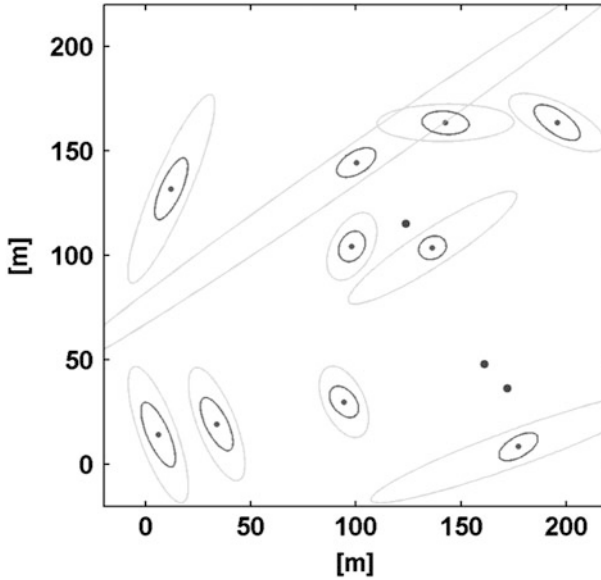
Fig. 8.15 Comparison of the positioning performance of non-cooperative (*light gray*) vs. cooperative and distributed (*bold gray*) positioning. The anchors are represented by big dots, the located mobile terminals by small dots

8.3.4 Conclusions

In 2014 in the UK a geo-location database is planned by OFCOM with cell sizes of $100 \times 100\,\text{m}^2$ [42]. London city would cover about 280 of such cells for 300,000 people each day. That means a single cell is responsible for more than 1,000 people. On the other side 1,000 people may offer a location information about their whereabouts and may even consider the 3rd dimension, which is in a city with high buildings another valuable degree to exploit. This chapter surveys that a much more accurate position performance especially indoors is possible. The different distributed cooperative location algorithms offer in very dense scenarios precise location of about the human level size, which is approximately 1 m. Especially in dense scenarios the positioning performance improves significantly jointly with the need for a more precise spectrum allocation for multiple users. Therefore, the achievable performance of the location algorithms supports the cognitive radio concept of sharing spectrum in a flexible way between all participating radio nodes.

References

1. Akyildiz, I.A., Won-Yeol, L., Chowdhury, K.: Spectrum management in cognitive radio and ad-hoc networks. IEEE Netw. **23**(4), 6–12 (2009a)
2. Chochliouros, I.P., Spiliopoulou, A.S.: Cognitive systems in the scope of modern European strategic priorities. In: Virvou, M., Nakamura, T. (eds.) Knowledge-Based Software Engineering, pp. 528–539. IOS Press, Amsterdam (2008)
3. Mitola, J., III: Cognitive radio for flexible multimedia communications, In: IEEE International Workshop Mobile Multimedia Communications'99 (MoMuC'99), pp. 3–10. San Diego, California (1999)
4. Mitola, J., III, Maguire, G.Q.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
5. Wyglynski, A., Nekovee, M., Hou, T.: Cognitive Radio Communication and Networks: Principle and Practice. Academic, London (2010)
6. International Telecommunication Union – Radiocommunication Sector (ITU-R): ITU-R Report SM.2152: Definitions of Software Defined Radio (SDR) and Cognitive Radio System (CRS), Geneva, Switzerland: ITU-R (2009)
7. Akyildiz, I.A., Won-Yeol, L., Chowdhury, K.: Crahns: cognitive radio ad-hoc networks. Ad Hoc Netw. **7**(5), 810–836 (2009b)
8. Le, B., Rondeau, W.T., Bostian, W.C.: Cognitive radio realities. Wirel. Commun. Mob. Comput. **7**(9), 1037–1048 (2007)
9. Thomas, R.W., DaSilva, L.A., MacKenzie, A.B.: Cognitive networks. In: First IEEE International Symposium on New frontiers in Dynamic Spectrum Access Networks 2005 (DySPAN 2005), pp. 352–360. IEEE, Baltimore, Maryland (2005)
10. Chochliouros, I.P., Spiliopoulou, A.S., Georgiadou, E., Belesioti, M., et al.: A model for autonomic network management in the scope of the future internet. In: Proceedings of the 48th FITCE International Congress, pp. 102–106. FITCE, Prague, Czech Republic (2009)
11. Mihailovic, A., Chochliouros, I.P., Kousaridas, A., Nguengang, G., et al.: Architectural principles for synergy of self-management and future internet evolutions. In: Proceedings of the ICT Mobile Summit 2009, pp. 1–8. IMC Ltd, Dublin, Santander (2009)
12. Raptis, T., Polychronopoulos, C., Kousaridas, A., Spapis, P., et al.: Technological enablers of cognition in self-manageable future internet elements. In: Proceedings of COGNITIVE-2009, pp. 499–504. IARIA, Athens/Glyfada (2009)
13. Ganesan, G., Li, Y.: Cooperative spectrum sensing in cognitive radio, part I: two user networks. IEEE Trans. Wirel. Commun. **6**(6), 2204–2212 (2007)
14. Ofcom: Digital dividend: cognitive access. Retrieved 28 June 2011, from http://stakeholders.ofcom.org.uk/consultations/cognitive/?a=0 (2009)
15. Granelli, F., Pawelczak, P., Prasad, R.V., Subbalakshmi, K.P., Chandramouli, R., Hoffmeyer, J.A., Berger, H.S.: Standardization and research in cognitive and dynamic spectrum access networks. IEEE Commun. Mag. **48**(1), 71–79 (2010)
16. Zhao, Q., Sadler, B.M.: A survey of dynamic spectrum access. IEEE Signal Process. Mag. **24**(3), 79–89 (2007)
17. Nekovee, M.: A survey of cognitive radio access to TV white spaces. Int. J. Digit. Multimed. Broadcast. **2010**, 1–12 (2010). (Hindawi Publishing Corporation)
18. ETSI (European Telecommunications Standards Institute): ETSI TR 102 802 V1.1.1 (2010-02): Reconfigurable Radio Systems (RRS); Cognitive Radio System Concept. ETSI, Sophia-Antipolis (2010, February)
19. ETSI (European Telecommunications Standards Institute):. ETSI TR 102 838 V1.1.1 (2009–10): Reconfigurable Radio Systems (RRS); Summary of Feasibility Studies and Potential Standardization Topics. ETSI, Sophia-Antipolis (2009, October)
20. Khalid, L., Anpalagan, A.: Emerging cognitive radio technology: principles, challenges and opportunities. Comput. Electr. Eng. **36**(2), 358–366 (2010)

21. Fortuna, C., Mohorcic, M.: Trends in the development of communication networks: cognitive networks. Comput. Netw. **53**(9), 1354–1375 (2009)
22. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009)
23. Popescu, A., Erman, D., Fiedler, M., Popescu, A., Kouvatsos, D.: A middleware framework for communication in cognitive radio networks. In: International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT), Moscow, pp. 1162–1171. IEEE (2010)
24. Institute of Electrical and Electronic Engineers (IEEE): IEEE Std 1900.4-2009: IEEE standard for architectural building blocks enabling network-device distributed decision making for optimized radio resource usage in heterogeneous wireless access networks. IEEE, New York (2009)
25. Sayed, A.H., Tarighat, A., Khajehnouri, N.: Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. Signal Process. Mag. **22**(4), 24–40 (2005). doi:10.1109/MSP.2005.1458275
26. Niculescu, D., Nath, B.: Ad hoc positioning system (APS) using AOA. In: Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003), vol. 3, pp. 1734–1743. IEEE Societies, San Francisco, CA. doi:10.1109/INFCOM.2003.1209196
27. Li, X., Pahlavan, K.: Super-resolution TOA estimation with diversity for indoor geolocation. IEEE Trans. Wirel. Commun. **3**, 224–234 (2004). doi:10.1109/TWC.2003.819035.
28. Chen, L.-H., Chen, G.-H., Jin, M.-H., Wu, E.H.-K.: A novel RSS-based indoor positioning algorithm using mobility prediction. In: 39th International Conference on Parallel Processing Workshops (ICPPW), San Diego, CA, pp. 549–553 (2010). doi:10.1109/ICPPW.2010.80
29. Le Dortz, N., Gain, F., Zetterberg, P.: WiFi fingerprint indoor positioning system using probability distribution comparison. In: 2012 IEEE International Conference on, Acoustics, Speech and Signal Processing (ICASSP), pp. 2301–2304 (2012). doi:10.1109/ICASSP.2012.6288374
30. Gunawan, M., Li, B., Gallagher, T., Dempster, A.G., Retscher, G.: A new method to generate and maintain a WiFi fingerprinting database automatically by using RFID. In: 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6. Sydney, NSW (2012). doi:10.1109/IPIN.2012.6418881
31. Brown, D.R., Dunn, D.B.: Classification schemes of positioning technologies for indoor navigation. In: 2011 Proceedings of IEEE Southeastcon, Nashville, TN, pp. 125–130 (2011). doi:10.1109/SECON.2011.5752919
32. WHERE2 Project. http://www.ict-where2.eu
33. Steinböck, G., Pedersen, T., Fleury, B.H., Wang, W., Raulefs, R.: Distance dependent model for the delay power spectrum of in-room reverberant channels. IEEE Trans. Antennas Propag. **61**(8), pp. 4327–4340 (2013)
34. WHERE2 Partners: Final: location information extraction (FP7-ICT-2009-4 WHERE2 Deliverable D2.5). http://www.ict-where2.eu/documents/Deliverables/Deliverable-D2.5.pdf (2013)
35. WHERE2 Partners: Final: self-learning positioning using inferred context information (FP7-ICT-2009-4 WHERE2 Deliverable D2.6). http://www.ict-where2.eu/documents/Deliverables/Deliverable-D2.6.pdf (2013)
36. WHERE2 Partners: Final: synergetic cooperative location and communications for dynamic heterogeneous networks (FP7-ICT-2009-4 WHERE2 Deliverable D2.4). http://www.ict-where2.eu/documents/Deliverables/Deliverable-D2.4.pdf (2013)
37. Mensing, C., Sand, S., Dammann, A.: Hybrid data fusion and tracking for positioning with GNSS and 3GPP-LTE. Int. J. Navig. Obs. (2010). doi:10.1155/2010/812945
38. Denis, B., Raulefs, R., Fleury, B.H., Uguen, B., Amiot, N., de Celis, L., Dominguez, J., Koldsgaard, M.B., Laaraiedh, M., Noureddine, H., Staudinger, E., Steinboeck, G.: Cooperative and heterogeneous indoor localization experiments. In: Proceedings of the IEEE International Conference on Communications (IEEE ICC'13), Budapest (2013)
39. Raulefs, R., Zhang, S., Mensing, C.: Bound-based spectrum allocation for cooperative positioning. Trans. Emerg. Telecommun. Technol. **24**(1), 69–83 (2013) Wiley. doi:10.1002/ett.2572

40. Zhang, S., Raulefs, R., Dammann, A., Sand, S.: System-level performance analysis for Bayesian cooperative positioning: from global to local. In: Proceedings of the Fourth IPIN International Conference – Indoor Positioning and Indoor Navigation (IPIN 2013), Montbéliard-Belfort (2013)
41. Zhang, S., Raulefs, R.: Improved particle filtering by exploring nomadic movements. In: 5th International Symposium on Communications, Control, and Signal Processing (ISCCSP 2012), Rome (2012)
42. OFCOM: TV white spaces: approach to coexsistence. http://stakeholders.ofcom.org.uk/binaries/consultations/white-space-coexistence/summary/white-spaces.pdf (2013)

Chapter 9
Challenges Towards a Cloud-RAN Realization

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

Abstract In order to sustain and ultimately succeed in the Future Internet (FI) era, wireless/mobile broadband networks will have to jointly satisfy complex context of operations as well as system requirements like QoE (Quality of Experience), energy and cost efficiency. Introduction of intelligence in the Cloud-RANs will lead towards this direction by providing the necessary decisions based on the received inputs. Cloud-RANs have the capabilities to adapt their network topology and resource allocation so as realize environmental-friendly and cost-efficient solutions by moving elements of the legacy RAN to cloud-based infrastructures. Therefore, through this chapter we will try to provide an indication on the elements of the approach as well as the identification of the benefits from such a concept.

9.1 Introduction

Wireless, mobile broadband networks are moving to an era where intelligent management of valuable resources (e.g., communication, computing, storage, energy) will become vital in order to maintain operational efficiency. At the same time wireless world is expected to be characterized by complex context of operation and multiple objectives which should be successfully met so as to effectively deploy powerful systems. Complex context of operations is considered to take into account the following main aspects:

- Increasing and variable (in terms of time and space) traffic demand (by a factor of at least 10×) [1], which is due to (i) the anticipated larger volume/diversity of

A. Georgakopoulos (✉) • D. Karvounas • V. Stavroulaki • K. Tsagkaris • P. Demestichas
University of Piraeus, Piraeus, Greece
e-mail: andgeorg@unipi.gr; dkarvoyn@unipi.gr; veras@unipi.gr; ktsagk@unipi.gr;
pdemest@unipi.gr
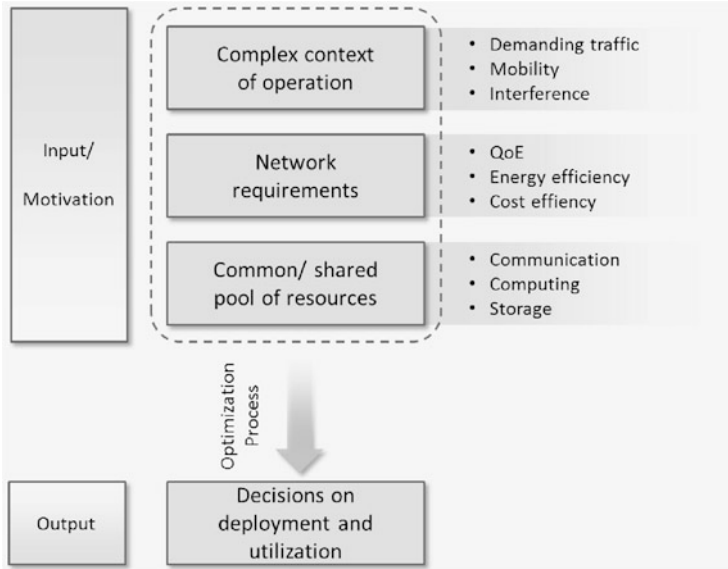
Fig. 9.1 Motivation

users that will want to use the wider set of offered applications and services; (ii) the anticipated large number of heterogeneous machines, also on account of the Internet of Things (IoT) concept [2].

- Various mobility levels due to the fact that the wireless world includes moving users which need to be served seamlessly;
- Interference due to the fact that heterogeneous deployments may include small cells which may interfere to other devices if not properly configured/positioned.

In addition, multiple requirements need to be met including QoE (Quality of Experience) satisfaction, operation of systems in an energy efficient manner, resource efficiency and cost efficiency. Cost-efficient application provision is also essential in order to lower the operational and capital expenditures, human intervention and total cost of ownership in general. The need of jointly addressing all these complex situations and requirements leads to the introduction of intelligence to the emerging wireless world.

Cloud-RAN (Radio Access Network) should also be enhanced towards this direction. Main concept of the Cloud-RAN is based on the fact that resources are available through cloud-based infrastructures. Therefore, it becomes possible to utilize common/shared resources; to select the volume/type of functional components that should be activated and to select the physical elements of the cloud, on which functionality should be deployed and their interconnections. Figure 9.1 provides an overview of the motivation and the expected output/decisions.

Main goal of this chapter is to describe the main concept of the Cloud-RAN, and to identify specific functionality which could then be enhanced through intelligence

and cognition. Consequently, decision-making is expected to be more efficient so as to unleash and fully exploit the features and capabilities of Cloud-RAN infrastructures.

The rest of the chapter is structured as follows. Section 9.2 presents dominant literature on the subject. Section 9.3 provides a description of the concept and introduces intelligence and cognition. Elements of the approach and main, expected benefits are also analyzed. Section 9.4 elaborates on key future considerations and the bonding of Cloud-RAN with virtualization and the emerging concept of network function virtualization (NFV). Finally, Sect. 9.5 discusses main conclusions of the work.

9.2 Related Work

An important aspect of the next generation wireless networks has to do with the cloud concept and the migration from a more static approach where dedicated, expensive hardware resources are deployed, into a more flexible utilization of resources. Specifically, the notion of the infrastructure as a service (IaaS) is described in [3] where computing power, storage and network connectivity is being virtualized and users have the ability to scale these resources on demand. Cloud computing can act as an enabler for the effective delivery of services to users through flexible and scalable resource virtualization. Furthermore, the cloud-based RAN concept assumes that RAN components can be virtualized and implemented in a centralized manner, using real-time cloud computing. The introduction of the Cloud-RAN is motivated by further reduction in total cost of ownership (TCO) and by the need of higher flexibility in terms of the network optimization [4–6]. Two possible Cloud-RAN architectures were identified in [4]. The first architecture is fully centralized while the second architecture is partially centralized and considers the centralized processing for higher layers only. Although the Cloud-RAN is still in early stages of development, possible structures for the realization of the Virtual Pool were also proposed. According to [7], the functions of a base station can be implemented in software using one or more software instances.

Additionally, main advantages and disadvantages of the novel, Cloud-RAN architecture are discussed by authors in [8]. Potential advantages are covering aspects like lowering deployment and operating costs of the radio access network, system throughput improvement, optimization of resource utilization, smooth coexistence of multiple standards and joint transmission and processing. On the other hand, potential disadvantages are focused according to the authors on deploying Cloud-RAN systems and constructing optical infrastructure in order to connect various cell sites to a centralized cloud infrastructure. Moreover, authors in [9] elaborated on the notion of RAN as a Service and they are identifying challenges and solutions e.g., split of the functionality executed at the base stations and at the Cloud-RAN etc. Also, authors in [10] elaborate on the concept of re-configurable backhaul in Cloud-RAN based small cell networks by proposing also a novel

architecture which considers decoupling of baseband unit processing from remote radio access units and user buffers, baseband selection and scheduling, reconfigurable switch and optical conversion and distribution. Experimental evaluation is also provided as a proof-of-concept.

The aforementioned related work provides a justification of the importance of the issue of Cloud-RANs by exploiting the potentials and advantages and trying to mitigate any obstacles that may be experienced. Framed in this context, this publication provides an advancement of the Cloud-RAN concept by proposing the introduction of intelligence and cognition in order to stress on the benefits from the deployment of Cloud-RANs compared to legacy RANs.

9.3   Intelligence and Cognition in Cloud-RANs

This section addresses the issue of intelligence and cognition in the Cloud-RANs. Main elements of the approach are discussed, in order to show the tasks which can be enhanced through intelligence. Introduction of intelligence and cognitive management is one of the main challenges in the emerging concept of Cloud-RANs, which will lead to more efficient decisions in terms of functionality allocation to physical resources, activation of appropriate shared elements etc.

In order to enable the making of intelligent and cognitive decisions, the network needs to gain knowledge through the acquisition of contextual information (i.e., the current status of the involved elements such as servers, base stations etc.), profiles and any related policies. The aforementioned parameters will create the input, according to which decisions will be made.

In addition, Cognitive Radio (CR) will also have a complementary role by providing knowledge features on the decision making. This means that decisions will be monitored with respect to their performance and a tuple of information will be stored, comprising:

  (i) Context, profiles, policies;
 (ii) Decision made;
(iii) Performance achieved

The available information will contribute to the knowledge process, which will facilitate to faster and more efficient decisions when similar conditions are met in the future. The subsection that follows provides an analysis of the elements of the approach.

9.3.1   Elements of the Approach

Figure 9.2 illustrates the overall concept of the Cloud-RAN approach compared to the legacy one. Specifically, the Cloud-RAN is based on the fact that essential
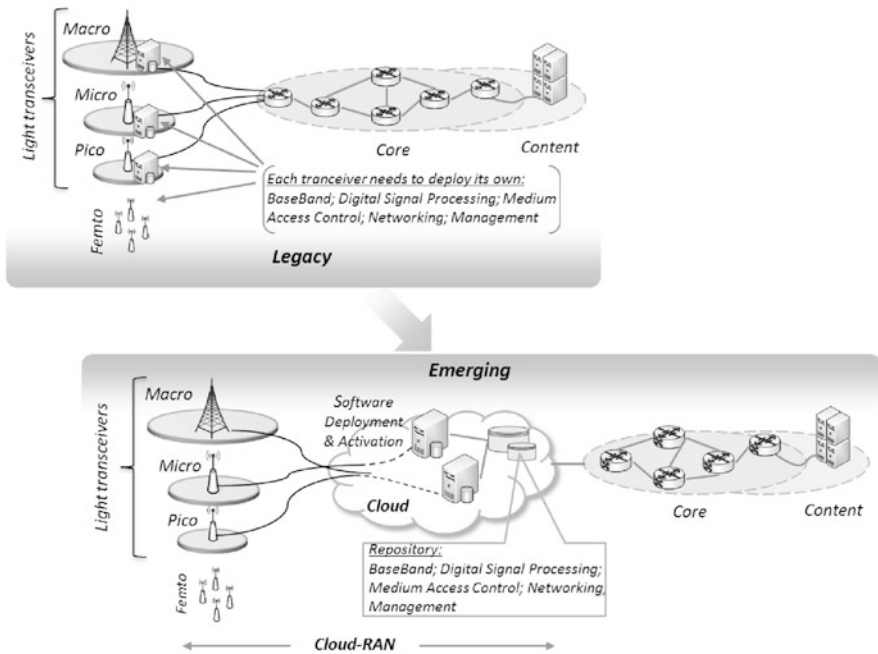
Fig. 9.2 Concept

functionality of each physical, infrastructure element can migrate to centralized repositories and processing machines located to the cloud. Consequently, it becomes possible to utilize specific computing, communication and storage resources to physical elements that actually need them and avoiding over-provisioning of resources to under-utilized physical elements.

Decisions related to the proper resource provisioning possess the need for the introduction of intelligence into the decision-making mechanisms. Through intelligence it will be possible to:

(i) Select physical elements/transceivers that should be involved in situation handling, their configuration (respective resource assignment), and the traffic distribution to the cells. This means that not all available elements are necessarily involved in the handling of a specific situation (e.g., capacity expansion of the infrastructure). As a result, intelligent mechanisms should be provided so as to proceed to the selection of the most appropriate elements and the distribution of traffic;

(ii) Activate the appropriate volume/type of functional/software components needed (beyond the transceiver-bound ones) for handling the situation. This decision resembles the previous one in terms of selection of elements which are not hardware at this time, but are functional/software components. The selected components will proceed to the handling of situations;

(iii) Allocate functional components to the physical elements. Framed in this context, this decision will designate the mapping of each functional component to a physical element;

(iv) Select the best connections between transceivers and physical elements, and the best interconnections between the physical elements. It could be noted also here that in order to realize reliable Cloud-RANs, installation of fiber optics between the base stations and servers need to take place so as to ensure high-speed links between these entities.

Through these decisions there is the derivation of the wireless network configuration that is appropriate for handling the situation faced. The aforementioned decisions are the main challenges of a Cloud-RAN infrastructure, since performance can be directly or indirectly affected.

In addition, through cognition it will be possible to provide more accurate and faster decisions based on acquired knowledge from previous configurations and similar contexts of operation. The exploitation of knowledge can further benefit Cloud-RANs by accumulating the characteristics of situations, on how they were handled (actions taken and outcome), as well as on the potentials of alternative solutions. So this knowledge can be used for: (i) making decisions faster (e.g., the inspection of a knowledge base and the identification of a relevant good case can be an alternative or a complementary measure/influence to complex optimization algorithms); and (ii) influencing the system operation.

9.3.2   Benefits from the Cloud-RAN

Benefits of Cloud-RANs can be manifold. According to [4], a list of benefits includes:

- Energy efficiency;
- Cost savings;
- Capacity improvement/Offloading;
- Adaptability to traffic fluctuations.

Flexibility is added to the network, since the centralized deployment of features such as "Baseband", "Digital Signal Processing", "Medium Access Control" and "Management functionality" gives the opportunity to the operators to allocate the available resources dynamically and more effectively, wherever they are needed, compared to legacy situations, by being able to select from common repositories. Flexibility, increases also the adaptability to traffic fluctuations, which is a vital feature of the Cloud-RAN, since it becomes possible to improve stability and better handling of the decisions related to resource allocation from common repositories to physical elements.

In general, "adaptability" means to pursue proper situation handling through the orientation and dedication of the necessary network resources, where they are
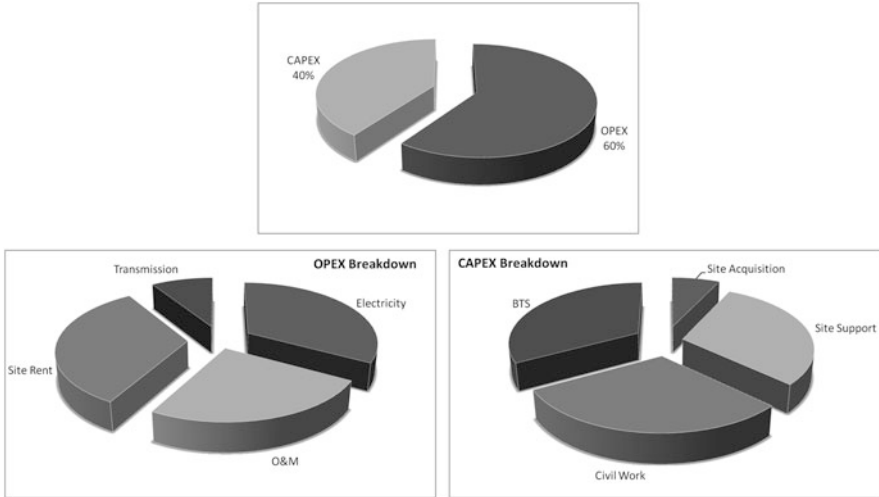
Fig. 9.3 Breakdown of total costs as reported by a mobile operator

needed and for the amount of time they are needed. To that respect, there can be use of the necessary subset of the overall network capabilities therefore, this feature is inline with the requirement to deploy few resources (over-provisioning can be avoided, since network resources can be re-assigned/oriented appropriately) and to operate few resources (use of the necessary subset of network resources implies that there can be releases/switch-offs of the resources that are redundant in the particular situation handling).

In turn, dynamic allocation implies that the adaptations of the network to the situation are conducted whenever required and possible, mainly in real-time but also at medium/longer time scales, and (by default) through dynamic/automated mechanisms; therefore, through this feature, optimality is constantly sought/preserved and also there is compliance with the requirement to automate.

Additionally, optimality can involve all the factors deriving from the requirements such as

  (i) The QoE levels offered (quality of the handling of each situation);
 (ii) The selection of an adequate and (in principle) minimal set of communication resources that will be involved (consumed) in the handling of each situation;
(iii) The realization of adaptations between situations;
(iv) The minimization of the energy consumption, e.g., by switching-off resources that are redundant in the particular situation handling.

Also, it is worth noting that as less processing machines are needed in order to handle resources of several infrastructure elements (compared to the legacy 1-to-1 mapping), this is expected to lead to reduced energy consumption as well as reduced capital expenditures (CapEx) and cost savings. The CapEx and OpEx (operating expenditures) breakdown is depicted in Fig. 9.3. According to [4], with
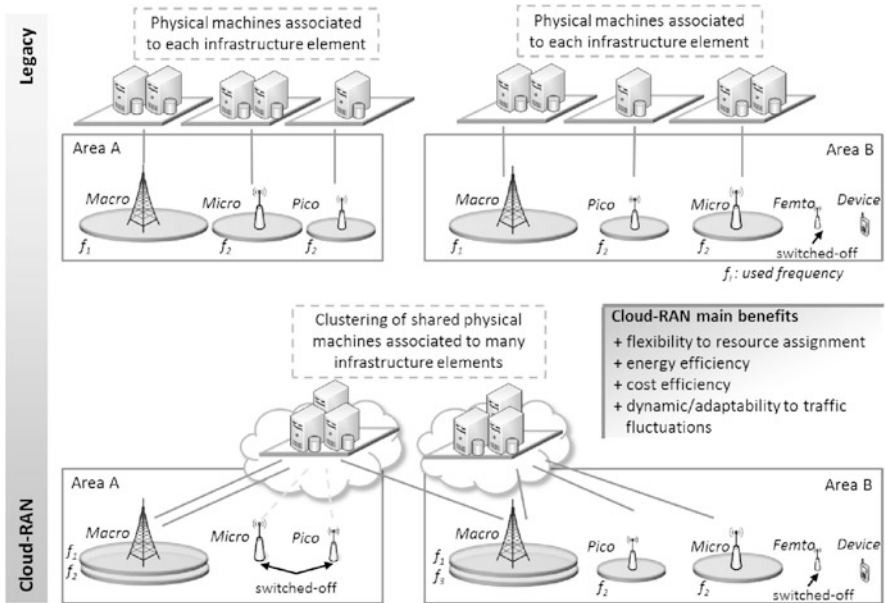
Fig. 9.4 Main benefits of Cloud-RAN

is based on data of a leading mobile operator with 700 million subscriber base and expanding, it is evident that CapEx is accounting almost for the 40 % of the total costs, while OpEx for almost 60 %. In this respect, when centrally-allocated resources are utilized in the network through Cloud-RANs, it is projected that can result CapEx savings of around 30 % and OpEx savings of more than 50 %. OpEx is benefited from the fact that need for electricity, O&M and site rental is reduced. In addition, CapEx is benefited from the fact that less equipment and civil work are needed, but at the same time network capabilities remain the same.

Figure 9.4 illustrates the main benefits of the Cloud-RAN approach compared to the legacy one. Specifically, in the legacy situation (as depicted in the upper part of the figure) the various base stations and access points are associated to dedicated infrastructure elements such as servers, baseband units etc. Also storage usually is not shared in legacy systems. Through the cloudification of resources (as depicted in the lower part of the figure), it is possible to proceed to clustering of shared physical infrastructure, so each base station and access point can have e.g., common storage facilities, common processing capabilities etc. The obvious benefit is cost efficiency (due to the acquisition of less but shared infrastructure elements), energy efficiency and reduced personnel costs for management and monitoring of infrastructure.

9.4 Future Considerations

The Cloud-RAN concept can be also supported through the introduction of virtualization. In general, virtualization is considered as a resource abstraction that allows resource sharing. This abstraction is usually implemented as a software layer that provides virtual interfaces similar to the real resource interface. For example in computer virtualization, the Virtual Machines (VMs) are provided with an interface (i.e. the hardware abstraction layer) which is similar to the real hardware interface (e.g. processor, memory, etc.). Therefore, each VM is under the impression that operates over the physical hardware. This technique spares energy and maintenance cost, but the most important feature is flexibility as each VM can have its own operating system, applications, configuration, etc. [11]. Similarly, network virtualization not only allows adding, removing and monitoring of virtual networks, but also it allows the migration of network elements and the configuration of resource allocation parameters. Therefore, network virtualization can provide the following benefits:

- Creation of multiple customised networks which can include virtual network elements (e.g. virtual router, virtual links, etc.) and they can be rapidly deployed and run in parallel;
- Flexible management as the link between the forwarding table and the physical hardware that implements the packet forwarding mechanism is broken [12]. Therefore the migration of a virtual network element from a physical hardware to another without ruining the logical/virtual network topology is possible;
- Real time control as resource allocation parameters can be configured for each virtual network element. Thus, resource allocation to each virtual network can adapt to the current network status;
- Monitoring as many variables that can be measured (e.g. bandwidth, delays, etc.). Furthermore an intrusion detection system is available which can detect malicious nodes. In such case, a virtual network element or even an entire virtual network can be removed.

An emerging, key enabler towards realization of virtualization in physical infrastructures is the concept of Network Function Virtualization (NFV) [13]. NFV is based on cloud computing principles and involves high volume servers which are capable of dynamically hosting the network functionality. Through the capability of combining the network functions of many physical elements to a subset of virtual machines, economies of scale are possible to be achieved in the IT industry. Additionally, NFV would eventually lead to reduced CapEx and reduced energy consumption, since operators may take advantage of the virtualization functionalities and proceed to the deployment of less infrastructure elements. It becomes also evident that essential intelligence is necessary in order to manage (activate, deploy, cease) functionalities and physical elements.

Figure 9.5 illustrates the overall concept of the Cloud-RAN approach enhanced with virtualization functionality. Specifically, the physical infrastructure is depicted
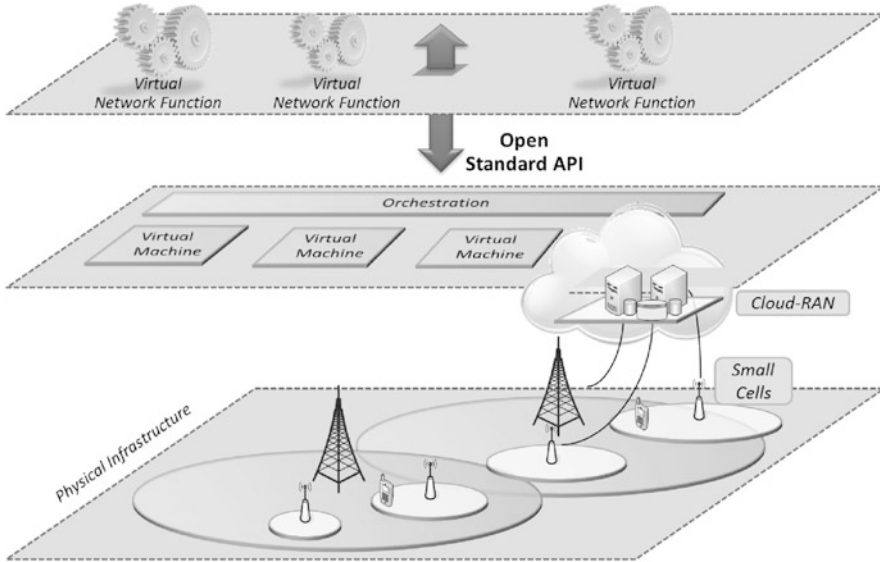
Fig. 9.5 Virtualization functionality

at the lower part of the figure. Elements of the physical infrastructure make use of shared resources in the Cloud-RAN, which can also interact with virtual machines (in a coordinated manner) in order to use virtualized network functions.

9.5 Conclusions

Through this work, an overview of the Cloud-RAN concept and its benefits was made. Also, the chapter identified the anticipated impact deriving from the introduction of intelligence and cognition to Cloud-RANs. Finally, an elaboration on the notion of virtualization was included in addition to the NFV concept description which can act as an enabler towards the Cloud-RAN realization. Cloud-RAN can be a promising solution which can lead to decreasing operational, capital and energy costs, due to the fact that each base station can share processing, storage, computing resources from shared infrastructure elements. As a requirement, installed fiber optics between the base stations and the shared servers can be provided, in order to ensure fast and reliable transmission of data. Operators have estimated the benefits of cloudification of RANs to be around 30 % for CapEx savings and around 50 % for OpEx savings due to reduced electricity, O&M and site rental. Because of the fact that the Cloud-RAN concept is rather new and not widely implemented by the mobile operators yet, future considerations of this work will include further experimentation of the concept, in order to validate new approaches related to Cloud-RANs.

References

1. Actix: Data growth will transform mobile infrastructure by 2015, available online at: http://www.emlwildfire.com/primages/acts-smallcells2015.pdf. Accessed May 2013
2. European Research Cluster on the Internet of Things: website: http://www.internet-of-things-research.eu/. Accessed May 2013
3. Yau, S., An, H.-G.: Software engineering meets services and cloud computing. IEEE Comput. **44**(10), 47–53 (2011)
4. China Mobile Research Institute: C-RAN – road towards green radio access network. White Paper (2011)
5. Lin, Y., Shao, L., Zhu, Z., Wang, Q., Sabhikhi, R-K.: Wireless network cloud: architecture and system requirements. IBM J. Res. Dev. **54**(1), 4:1–4:12 (2010)
6. Huawei: Cloud RAN introduction. In: Proceedings of 4th CJK International Workshop, Bundang (2011). Available: edu.tta.or.kr/sub3/down.php?No=88&file=3-2%20Cloud-RAN.pdf. Accessed November 2013
7. Zhu, Z., Gupta, P., Wang, Q., Kalyanaraman, S., Lin, Y., Franke, H.: Virtual base station pool: towards a wireless network cloud for radio access network. In: Proceedings of 8th ACM International Conference on Computing Frontiers, Bertinoro (2010)
8. Hadzialic, M., Dosenovic, B., Dzaferagic, M., Musovic, J.: Cloud-RAN: innovative radio access network architecture. In: Proceedings of International Symposium (ELMAR), Zadar, pp. 115–120 (2013)
9. Sabella, D., Rost, P., Sheng, Y., Pateromichelakis, E., Salim, U., Guitton-Ouahou, P., Di-Girolamos, M., Giuliani, G.: RAN as a service: challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network. In: Proceedings of Future Network and Mobile Summit (FNMS), Lisbon, pp. 1–8 (2013)
10. Liu, C., Sundaresan, K., Jiang, M., Rangarajan, S., Chang, G.-K.: The case for re-configurable backhaul in Cloud-RAN based small cell networks. In: Proceedings of IEEE INFOCOM, Turin, pp. 1124–1132 (2013)
11. Fernandes, N., Moreira, M., Moraes, I., Ferraz, L., Couto, R., Carvalho, H., Campista, M., Costa, L., Duarte, O.: Virtual networks: isolation, performance, and trends. Ann. Telecommun. **66**(5–6), 339–355 (2010)
12. Wang, Y., Keller, E., Biskeborn, B., van der Merwe, J., Rexford, J.: Virtual routers on the move: live router migration as a network-management primitive. In: Proceedings of ACM SIGCOMM, Seattle, pp. 231–242 (2008)
13. ETSI Industry Specification Group: Network function virtualisation. http://portal.etsi.org/tb.aspx?tbid=789&SubTB=789,795,796,801,800,798,799,797,802 (2013). Accessed May 2014

Chapter 10
A Regulatory Perspective on Cognitive Radio and Spectrum Sharing

Linda Doyle and Tim Forde

Abstract The purpose of this chapter is to give a regulatory perspective on cognitive radio. Cognitive radio can be seen as a natural part of the roadmap for advanced communication systems and from this standpoint can be dealt with within the context of normal regulations. However one of the key and unique advantages of cognitive radio is that it is an enabler of spectrum sharing in its many forms. Hence the main part of this chapter is devoted to different regulations in spectrum sharing and the implications for cognitive radio. It looks at regulations which are in existence as well as emerging regulations. The chapter also provides a generic framework in which to place different sharing regimes.

10.1 Introduction

This chapter focuses on different policy issues that touch on cognitive radio. The purpose of the chapter is to frame the state-of-play of the cognitive radio regulatory world and to provide perspectives on current progress as well as the outlook for the future.

There are two broad lines along which the chapter can be framed. Firstly, a cognitive radio can simply be seen as an advanced form of radio and the *normal* regulations relevant to future communication systems can apply. Using the terminology developed in Virginia Tech that describes a cognitive radio as device that has a set of 'meters' which feed into a cognitive engine that outputs the settings for the 'knobs' on the radio, we can unpack this comment a little further. The meters are the various different sensing and observation mechanisms, the knobs correspond to the parameters of the radio system that can be configured. Consider now an Long

L. Doyle (✉) • T. Forde
CTVR, Trinity College, University of Dublin, Co. Dublin, Ireland
e-mail: ledoyle@tcd.ie; fordeti@tcd.ie

Term Evolution – Advanced (LTE-A) system. The reality is that the number of parameters (knobs) which can be set in such a system is vast. Variable parameters include numbers of antennas in use, multiple-input, multiple-output (MIMO) mode, modulation, power, frequency, scheduling options, number of frequency bands that are aggregated, spatial configuration of primary and serving cells associated with the aggregated spectrum among many others. It seems infeasible that the true potential of LTE-A networks can be achieved without some advanced decision-making and optimization based on contextual information. Hence it is possible to argue that the realization of LTE-A systems, whether explicitly labelled as such or not, will need some elements of cognition to function. Context will have to be determined through observation of the environment (meters) and the many parameters (knobs) configured (with the help of a cognitive engine) to suit the context. Taken from this perspective no specific regulations are needed for cognitive radio as 'cognition' becomes a natural part of the roadmap for future mobile communication systems.

Secondly, and more commonly, cognitive radio regulations can be seen in terms of the key feature which is enabled through cognitive radio namely *sharing*. The desire to share spectrum has been driven largely by the increasing demand for more of it. There are numerous graphs forecasting the explosive growth of future mobile data.[1] The term 'spectrum crunch' is used to re-emphasise the scarcity of the resource and a plethora of measurements exist to highlight the fact that assigned spectrum is not always used efficiently and effectively by the licensee. A second and also very powerful motivation for sharing comes from a more political 'open spectrum' perspective. This mindset advocates for making access to spectrum freer and easier with a belief that much like in the ISM band new forms of innovation will follow and most importantly opportunity for new players, including smaller entities, will emerge. Sharing is seen as a means of gaining access to a resource that is central to any wireless communication system. Irrespective of the motivation, cognitive radio, while meaning many things to many people, has always been seen as a key enabling technology for sharing. The meters by which the radio observes its environment are no longer confined to spectrum sensing and include of course database-driven approaches and the knobs increasingly include advanced configuration capabilities that control, diminish or align interference in a manner that allows different systems to coexist. Spectrum sharing can happen in many dimensions. It can happen spatially and temporally. There can be hierarchies involved. It can involve complex rules or light touch regulation. In this chapter we focus mainly on this second line of thought and hence *we provide a regulatory perspective on cognitive radio as enabler of sharing*.

The chapter is organised as follows. Section 10.2 provides a means of contextualising regulatory polices and rules both to emphasise what has taken place to date as well as to create a framework in which to understand the regulatory issues discussed in the remainder of the chapter. Section 10.3 concentrates on the TV

---

[1]http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

White Space as a first and key area of application for cognitive radio. Section 10.4 goes on to take a much wider view of spectrum sharing and the kinds of regulations which may emerge over the coming years. Section 10.5 summarises some of the activities that are carried out by the IEEE Standards Association which includes a focus on definitions and terms as well as on other issues that are relevant in the context of sharing. Section 10.6 then discusses how sharing can be systematised in existing technologies as well as looks at sharing versus clearing. Section 1.5 draws the chapter to a close.

10.2   A Means of Contextualising the Regulatory Perspective

It may seem a trivial question but the question 'what is spectrum?' is one that underpins any understanding of rules and regulations for spectrum management and cognitive radio. While we are aware that a number of authors have created general frameworks for thinking about spectrum [28] we look to more economic terms for this purpose. This approach builds on that taken in a review paper commissioned by the Australian Communication and Media Authority (ACMA) policy in 2007 on the economics of spectrum management [16]. Basic microeconomic theory ranks all publicly and commercially available resources in a given society along two criteria: (i) excludability (or exclusiveness), which define property rights and access; and (ii) rivalry, which examines simultaneous use and depletion. A good or service is said to be excludable when it is possible to prevent people who have not paid for it from having access to it, and non-excludable when it is not possible to do so. A rival good is a good whose consumption by one consumer prevents simultaneous consumption by other consumers. On the other hand a non-rival good can be consumed simultaneously by many without the experience of any one user being detrimentally affected. A simple table of giving example of commodities which fall under the different headings if shown in Table 10.1.

Spectrum does not fall into any one class and can actually fall along the continuum of excludable/non-excludable and rival/non-rival. Figure 10.1 recasts the characteristics in Table 10.1 as a continuum to emphasise the point. If we look at how we consume spectrum today under the two headings of licensed and unlicensed spectrum we can place the traditional exclusive license in the left hand corner as shown in Fig. 10.1. Everyone except the licensee is excluded and the spectrum is rival because licensed communication systems are deployed in a manner so as not to tolerate any other system within the same band. The traditional unlicensed management of the ISM band falls somewhere along the right hand side of the diagram as shown in Fig. 10.1. Obviously it is non-excludable and anyone can use the spectrum. There is some amount of opinion involved in exactly where it might lie on the rival/non-rival trajectory. It is non-rival because many systems can coexist in the same band. However there is some theoretical limit and for example too many uncoordinated WiFi systems in the same location may eventually cause problems for each other – i.e. some form of a *Tragedy of the Commons* occurs [17].

Table 10.1 Classification of goods and services

Characteristics	Excludable	Non-excludable
Rival	Private good: cars, icecream	Commons: fish stock, open land
Non-rival	Club good: golf club, cinema	Public good: lighthouse, road sign
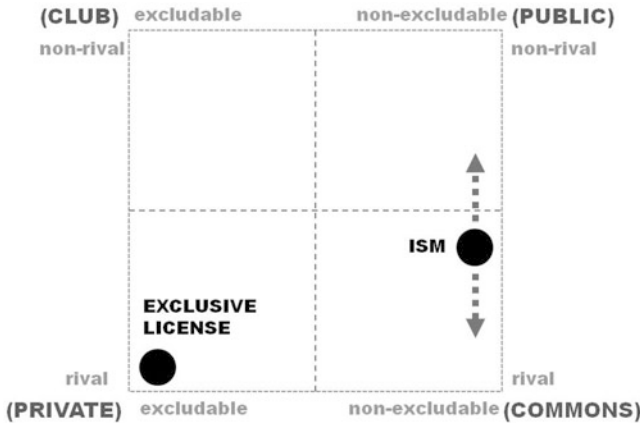
Fig. 10.1 The consumption of spectrum lies on a continuum, examples of specific modes of consumption are shown

The purpose of presenting the diagram in Fig. 10.1 is not to argue exactly where a licensing regime falls but to bring out two key insights. The first insight is that we can consume spectrum as any type of good we want by selecting the appropriate combination of spectrum access rules/policies and supporting technologies. If we use the term spectrum management framework to mean 'set of rules or policies for accessing the spectrum' we can see that different frameworks allow us to be more or less excludable. Frameworks that range from exclusive license scenarios, to scenarios in which a group of spectrum licensees (e.g. a group of mobile network operators or carriers) pool spectrum and share among themselves, or licensed shared access schemes that allow a chosen set of systems to share with incumbents to more open dynamic spectrum access regimes which allow any system willing to comply with a set of rules to consume the spectrum, steadily increase the members of the club making spectrum less and less excludable (i.e. move along the continuum from left to right). On the technology side advances in interference mitigation techniques, interference alignment techniques, advanced antenna techniques, database regimes, dynamic spectrum auction mechanisms, sensing techniques, techniques for shaping out-of-band emissions all contribute to making the consumption of spectrum less and less rival (i.e. move along the continuum from bottom to top). In other words they enable more and more systems to coexist or to share spectrum whether temporally or spatially or in a hierarchical or non-hierarchical fashion. Hence it is the combination of the rules/polices of the spectrum management framework
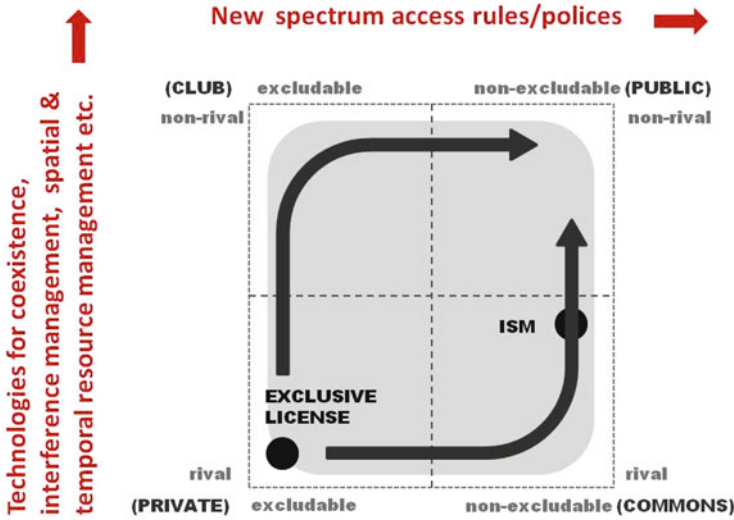
Fig. 10.2 We consume spectrum in a way that exploits a limited part of the continuum even though new rules and emerging technologies would allow us to have more options

and the technical mechanisms used within that framework that define a specific mode of spectrum consumption and place it at a given point on the excludable/non-excludable and rival/non-rival continuum. Therefore we can, in principle, consume spectrum any way we wish. This concept is captured in Fig. 10.2.

The second insight is that, in the main, the current modes of spectrum consumption occupy very little of the excludable/non-excludable and rival/non-rival continuum. In other words the true potential of emerging technologies and new spectrum management approaches are not yet exploited. As seen from the greyed out area of Fig. 10.2, this unexploited space gestures towards the increasingly includable rather than the excludable and the simultaneous/shared consumption rather than the rival. Expressed in a different way, this unexploited space is largely about new modes of sharing that can be unleashed with the help of cognitive technologies and appropriate spectrum management frameworks. The unexploited space is the focus of the chapter. It is the space within which new rules and policies are emerging that can leverage the capabilities of cognitive communication systems.

10.2.1   The Makers of the Rules

To end this section on contextualising the regulatory perspective, it is worth briefly commenting on the bodies and organisations which drive the regulatory policies and rules. Overall there are three tiers at which regulations and policies are made, namely international, regional and national.

The International Telecommunication Union (ITU), operates at international level and is a specialized agency of the United Nations that is responsible for issues that concern information and communication technologies. The ITU coordinates the shared global use of the radio spectrum. Membership of ITU is open to governments, which may join the Union as Member States, as well as to private organizations like carriers, equipment manufacturers, funding bodies, research and development organizations and international and regional telecommunication organizations, which can join ITU as non-voting Sector Members.

At regional level there are different bodies are at work depending on location. In Europe, for example, the European Union issues directives to member states, including directives on spectrum policy. There are two other key bodies working at the European regional Level. The European Telecommunications Standards Institute (ETSI) is an independent, non-profit, standardization organization in the telecommunications industry (equipment makers and network operators) in Europe, with worldwide projection. The European Conference of Postal and Telecommunications Administrations (CEPT) is a coordinating body for European state telecommunications and postal organizations. Within CEPT, the Electronic Communications Committee (ECC) is responsible for radio communications and telecommunications matters. The Committee for ITU Policy (Com-ITU) is responsible for organising the co-ordination of CEPT actions for the preparation for, and during the course of, the ITU activities. As a non European example Comisión Interamericana de Telecomunicaciones (Inter-American Telecommunication Commission – CITEL) is the Inter-American Telecommunication Commission.

At country level the national regulatory authority (NRA) is in charge. The NRA can comprise one or a number of bodies depending on the country. For example in the USA, the Federal Communications Commission (FCC) is an independent agency of the United States government responsible directly to Congress. It is charged with regulating interstate and international communications by radio, television, wire, satellite, and cable. The National Telecommunications and Information Administration (NTIA) on the other hand manages Federal spectrum.

Outside of these three tiers there are other bodies in operation that feed into regulations and standards. The Institute of Electrical and Electronics Engineers Standards Association (IEEE-SA) is an organization within IEEE that develops global standards in a broad range of industries, including information technology, and telecommunications. The IEEE-SA is not a body formally authorized by any government, but is considered to be an internationally active community.

The reality is that there are different activities at international, regional and national level as well as within bodies such as the IEEE Standards Association in which work on standards and policies for cognitive radio and spectrum sharing are taking place. There is no one systematic way to step through the work other than through creating a list. Rather than do this, the remainder of the chapter draws different material together under specific topics and mentions the different bodies where appropriate.

10.3   The Supposedly Low Hanging Fruit: The TV White Spaces

In practice we are still some way off realising the technical, economic and regulatory models and tools necessary to accomplish sophisticated forms sharing. But we have certainly begun that journey, albeit in a hesitant and uncertain manner.

The TV White Spaces are an interesting spectrum opportunity to look at as they indicate just how the collective of engineers, regulators and entrepreneurs have taken to, and shaped, a new spectrum paradigm. TV White Space has been to the fore in manifesting the intersection of Cognitive Radio, Spectrum Sharing an Dynamic Spectrum Access in tangible form.

The FCC started addressing TV White Spaces back in a Notice of Inquiry published in 2002 [22] in anticipation of the US Digital Switch-Over. In simple terms, the transition from analogue terrestrial TV systems to digital terrestrial TV (DTT) systems created an opportunity for new kinds of radio services and technologies to share the ultra high frequency (UHF) spectrum which was formerly exclusively allocated to broadcast TV. The digital TV signals are more spectrum efficient than their analogue predecessors and they are also more robust to interference. This has allowed regulators to clear the upper frequencies completely of TV services, re-allocate this spectrum for International Mobile Telecommunications (IMT) use and sell it off at auction for 4G services.

The high-power, high-tower topology of the DTT network now operates in the lower frequencies of the UHF spectrum. The arrangement of such a network results in a geographically interleaved frequency plan which leaves large pockets of unused UHF spectrum. The bandwidth and centre frequency of these spectrum guard-bands vary from location to location as dictated by the DTT topology and frequency plan. It is this spectrum which is the *TV White Space*. And it is this under-utilised space that the TV networks are expected to share.

Sharing can be a contentious activity. Unless all the participants understand the rules or the etiquette and think they are fair, whatever fair may be, then sharing a resource will be difficult. The participants in the TV White Space sharing model are the TV broadcasters, the wireless microphones and the newcomers, the cognitive TV White Space radios. However, each of these participants are quite unequal in terms of the technologies, services, business models they use and spectrum rights that they enjoy.

While the broadcasters are the primary, protected users of this spectrum the rights that they hold to use this spectrum vary; they have largely accumulated rights over time. In large part, their licenses to this spectrum, such as they are, are non-tradeable and they cannot unilaterally use the spectrum for other technologies and services. The other protected parties are the wireless microphone users, sometimes known as Programme Making and Special Events (PMSE) users. This class of users have been sitting alongside the TV broadcasters for some time. They utilise low-power, narrow-band signals in the unused channels and have been doing this during both

the analogue and digital phases of broadcast TV. However, the PMSE users operate on a non-interference, non-protected basis with respect to the TV broadcasters, they are second class users of this UHF spectrum. Wireless microphones can be thought of as the original TV White Space users.

That being said, the TV White Space could be considered low-hanging fruit in the spectrum sharing and secondary cognitive radio exploitation space for a number of reasons. The amount of TV White Space spectrum that is potentially available can be quite large. In Europe, where the 470–790 MHz band is set aside for UHF TV every country has different actual demands on that spectrum for its DTT network. These demands are driven by the topography of the country, which will dictate how the network's frequency plan will have to be adjusted for hilly or mountainous terrains, and also by the actual bandwidth demands of the TV operators; how many DTT multiplexes need to be put on air. The more DTT multiplexes that must be put on air, the less White Space that will be left over. Once this planning is done and the network has been rolled out and frequency-tuned then the TV White Space has one very enviable trait: it is a very stable resource. DTT networks are not subjected to frequent retunings, new in-fill transmitters are rarely required. As such, it is quite easy to calculate what spectrum will be available for sharing for a considerable period of time. While the PMSE users must also be considered, and these users can be less predictable, they still only use a fraction of the TV White Space.

So, from the perspective of a cognitive radio promoter, be it an engineer, entrepreneur or policy-maker, the TV White Spaces would seem to represent a relatively simple initial real-world deployment challenge.

10.3.1   TV White Space Regulations in General

In spite of being a seemingly attractive place in which to regulate for cognitive radios the process of formulating policy for the TV White Spaces has not been straightforward. This has been influenced by a number of factors; the incumbents, the technologies underlying cognitive radio and the business models that support TV White Space radios.

The initial, and possibly naive, suggestion by the FCC was that it might be possible for opportunistic cognitive radios to sense out these spectrum holes and then occupy the spectrum. It asked if modifications to the Part 15 rules used to govern unlicensed devices in the ISM bands would suffice. The answer to that consultation was resoundingly negative. There was considerable push-back from the incumbents, namely the TV broadcasters and the PMSE lobby.

There was also the problem, which still lingers in current regulatory discussions, that the details of putative cognitive radios were still ill-defined. The gap between a simple adaptive radio and a possibly learning, self-configuring, unpredictable, artificial-intelligence-aided radio was too much for the operators of traditional radio services to bridge. It was basically a trust-gap. Various descriptions and

definitions had been put forward by regulators [4, 5, 23, 24, 31], academics [20] and standardisation groups such as IEEE 1900.1. More details on the IEEE P1900 can be found in Sect. 10.5. The definitions that were put forward were either too vague or too specific to an application domain. Furthermore there was a coniïňĆation of definitions; cognitive radio, software defined radio (SDR), opportunistic radio, dynamic spectrum access (DSA) radio were often used interchangeably. While a DSA application may use both SDR and cognitive radio to effect its operation, a SDR is not necessarily a DSA radio and a cognitive radio is not necessarily a SDR. So, from the incumbents' perspective we had an ill-defined new radio looking to use its spectrum under a completely new regulatory paradigm.

But the FCC had also postulated that a database could be used to allow access to the TV White Space. Such a database would have of registry of locations in which opportunistic radios could use the TV White Spaces. The lobbyists who rejected the sensing-only cognitive radio model latched onto the database-controlled model, an idea which was at odds with the concept of the independent cognitive radio. From the cognitive radio perspective, the database approach was initially seen to kill all of the innovative aspects of CR. Much of the research in the area of cognitive radio and DSA has been in the sensing domain which was now being all but ruled out as a feasible or necessary approach for TV White Spaces. However, as we will discuss in this and subsequent sections, the database turned out to be particularly enabling tool for broader spectrum sharing and cognitive radio developments.

So, over a number of years of iterative consultation by both the USA under the FCC and the UK under Ofcom, the database-driven approach to tiered shared access to the TV Band developed. The main element of novelty in this regulatory framework is the dependence on an electronic repository, the TV White Space Database, to control secondary user access to a band that has incumbent operators. Essentially, these databases would describe what spectrum was occupied by the TV broadcasters by modelling the frequency use of the DTT networks. Based on DTT receiver criteria it is possible to calculate the area around a given DTT transmitter where TV receivers should be able to receive the DTT signal. That area, for that UHF channel, then becomes protected. This is repeated for every DTT transmitter at every frequency it broadcasts. It is then necessary to calculate in what locations, on which frequencies and at what transmission power limit it would be safe to allow a cognitive radio to transmit such that its transmissions would not cause interference to receivers in the protected areas around DTT transmitters. Given how stable the TV networks are, the generation of data for a TV White Space Database is a relatively large but trivial task.

The emerging consensus is that the master device or basestation in an opportunistic TV White Space network would access the database, providing details about itself including its location. The database would then return a list of TV channels on which the TV White Space network could operate, indicating an upper maximum power limit. The TV White Space basestation would then communicate with the end user devices so that they would only use the channels it received permission for.

10.3.2 A Global Summary of the State of Play

It is no surprise that USA was the first country to adopt comprehensive regulations for the TV White Spaces. It finalised its rules in 2010 [11] and tweaked them in 2012 [12]. The rules adopted by the FCC are almost entirely transmitter-centric. They dictate locations, maximum antenna heights (30 m above ground for a Fixed device), maximum transmission powers and strict out-of-band emission limits for the new TV Band devices. The 2012 rules relaxed some of the height and power limitations so that Fixed basestations could serve larger areas while still protecting DTT receivers from interference. Such Fixed devices must contact the database on a daily basis to obtain an updated list of available White Space TV channels. These basestations can transmit at up to 1 W with an antenna providing an additional 6 dBi of gain, i.e. 4 W EIRP. Two types of portable user devices are allowed. So-called Mode II portable devices can operate either by accessing the database and giving their location or under the control of a Fixed basestation. So-called Mode I devices are controlled by either the Fixed basestation or by a Mode II portable device that has access to a database. In both cases, the portable devices are limited to 100 mW EIRP transmit power, or 40 mW EIRP if on a channel adjacent to an active TV channel.

Up until the final rules were adopted in 2010, the FCC had seemed intent imposing sensing conditions on the TV Band devices. It *had* suggested that all devices would have to be capable of sensing TV and wireless microphone signals at a level of -114 dBm with reference to an omni-directional antenna with 0 dBi gain. For a Fixed device, they suggested that the sensing antenna would have to be at least 10 m above ground so that it would not blocked from receiving signals originating from incumbent services. Devices were to be required to do such sensing on the channel they were occupying at least once every minute, to vacate a channel within 2 s of sensing an incumbent and to remove the channel from the list that the database had given them. These rules pleased neither side of the debate.

On the one side the likes of Dell, Motorola, the WiFi Alliance argued that the threshold should be relaxed to -107 dBm arguing that the more stringent level would lead to many false detections, whereas on the other side wireless microphone vendors such as Shure and the Society of Broadcast Engineers argued that it be made more strict at -126 dBm. The Motorola, Google, Carlson Wireless side also argued that the 10 m minimum height requirement be lowered to 3 m. And in turn Shure asked that the sensing interval be lowered to 10 s.

In the end, the FCC decided to entirely eliminate any requirements for sensing from the final rules. There were a number of reasons for their decision. Firstly, the FCC tests all equipment prior to certification and its tests of initial TV Band equipment had noticed the 'limitations of spectrum sensing' in their current form. But the FCC did not prejudice the development of better techniques in the future. Indeed, they note that there is nothing to stop manufacturers from including sensing on a voluntary basis to improve the quality of service experience by their TV Band devices. Secondly, the FCC found that the protection afforded by the database

for TV signals was adequate by itself. Finally, the need for sensing of wireless microphones was rejected on two counts. Wireless microphones in the USA operate under either Part 74 or Part 15 rules of Title 47 of the US Code of Regulations. The FCC has already excluded TV White Space devices from 2 channels in every market as a kind of safe harbour for any wireless microphones. These 2 channels could host up to 16 wireless microphones. Part 74 rules enable certain types of professional users to license their wireless microphones and this enables them to register their additional devices, assuming that they are already using the 2 reserved channels. These are then added to the TV White Space Database which affords them protection from cognitive TV Band devices and removes the need for sensing to protect them. The FCC is imposing constraints around this mechanism to mitigate against abuse of the process whereby wireless microphone users may attempt to squat in channels, e.g. there will be a 30 day notification period and a minimum use of 6–8 microphones per channel. Any other wireless microphones in the TV Band which have not been registered under Part 74 rules operate under Part 15 rules which cover unlicensed devices, the same rules governing the ISM band. Normally, Part 15 devices operate on a non-interference, non-protected basis. As such the FCC does not consider that they deserve any special protection from other cognitive TV White Space devices.

However, given that a large number of Part 15 wireless microphones are in use by large professional organisations the FCC has created a *register* for *unlicensed* wireless microphones operating at *large* venues. Such registrations will then be afforded the same protections as Part 74 wireless microphones. The Commission stipulate that applications will only be considered for venues such as Broadway theatres, Madison Square Garden, the Kennedy Centre in Washington DC, etc. Venues such as courthouses and convention centres would not normally qualify. These ad hoc responses to the needs of various incumbents do not necessarily lead to consistent regulation, or regulations that are easily harmonised or adopted by other jurisdictions.

The FCC chose not to operate the TV White Space Database itself. Rather, it decided to allow market forces to operate and has given permission to a limited number of operators. Among these database operators are Spectrum Bridge LLC, Telcordia and Google. The FCC will test the databases prior to approval and each one should give exactly the same basic response for a given location queried. The rationale for allowing multiple database operators is that it could stimulate service competition among them. Databases will be permitted to charge a fee to a Fixed TV Band device querying them for a list of available channels. However, the databases are only required to give a basic binary response, i.e. a list of channels that are free or not free. It would be open to a database operator to provide value-added services such as indicating the quality of the channels or taking the users device characteristics into account and indicating preferred channels.

However, the database is not a flawless spectrum-sharing tool. It is reliant on the propagation models that are used to calculate DTT service areas and it is also reliant on the protection criteria assumed for the victim-interferer analyses that ultimately determine how much White Space will be available. Indeed, Adaptrum LLC, a US

company developing TV White Spaces devices, challenged FCC's continued use of its curves which, they argue, are more suited to flat or moderate terrain, but which fail more rugged or mountainous terrain [11]. The model is based on 1966 state-of-the-art calculations which were appropriate to the manual calculations of the time. Instead, Adaptrum argued that the FCC should also allow the use of the Individual Location Longley-Rice (ILLR) model which the Commission *already* use for the predictions of their DTT coverage. The message to take away here is that there is no single answer as to what the correct numbers that go into the database are. Yet these numbers will determine just what spectrum is made available for sharing. The database is not necessarily neutral.

Given that there will be multiple databases within the USA and anticipating that there will be TV White Space available globally, there has also been an effort to regulate or at least voluntarily standardise how cognitive radios would interact with databases operated by different companies. Consequently the Internet Engineering Task Force (IETF) has supported the development of an open standard, the Protocol to Access White-Space (PAWS) Databases. This protocol would standardise the query language that would be used by TV Band devices and TV White Space databases, thus improving the interoperability of devices and databases on a global basis.

Aside from the FCC, Ofcom in the UK has been the most proactive regulator of TV White Space in Europe. It has progressed issues far ahead of its European colleagues and commences pilot trials of their database in 2013 for a 1 year period prior to any authorisation of the system.[2] The Ofcom approach is largely similar although FCC approach but uses a much more sophisticated model to generate a pixelated database which allows for much more accurate and localised determinations of acceptable power limits for TV Band devices. Ofcom have indicated that they will harmonise their system with any harmonised European rules, should they come to pass.

However progress is still slow at a European level. The FP7 CogEU[3] project took a systematic look at TVWS issues across Europe and the picture remains highly fragmented. Different countries have different concerns and different levels of capabilities (such as appropriate data to fill a database). In addition it could be argued that the mobile operators remain at best disinterested and at worst opposed to the use of the TVWS fearing it will undermine the value of their systems.

Nonetheless, the CEPT SE 43 (White Spaces – cognitive radio systems) Working Group has progressed technical analyses of the requirements for TV White Spaces. In 2011 they issued their first draft report [8] and they completed their work at the end of 2012 with the publication of two final reports recommending technical parameters for both radio devices [9] and the TV White Space database [10]. These reports now await political actions and decisions.

---

[2]http://stakeholders.ofcom.org.uk/binaries/consultations/geolocation/summary/geolocation.pdf
[3]http://www.ict-cogeu.eu

10.3.3 The Outlook

On the one hand the outlook for TVWS can seem promising. Over the past couple of years trials of TVWS have become popular. Some of the main early trials included the USA-based Wilmington trials and the UK-based Cambridge trials followed by trials in Singapore and Finland. More recently countries in Africa have become very involved and many feel that Africa represents and excellent opportunity for TVWS. Current trials indicate that TV White Space it will enable connectivity and coverage in areas that are hard to reach, but it will not deliver capacity that Western markets seem to require. A single channel is able to deliver about 10 Mbps over about 8 km using the FCC regulations for which the trial radios have been developed. By mid 2013, vendors such as Adaptrum LLC and Neul had developed commercial radio products for the TV White Spaces but there were no commercial radios capable of bonding channels together. The Neul equipment is targeted to M2M applications and is based on an open standard *Weightless*.[4]

There is also a more significant threat to the TV White Space and that is the possibility that these White Spaces disappear. At World Radio Congress (WRC) 2012 Middle Eastern and African countries put pressure on the ITU to allocate 694–790 MHz to mobile services and hence bring with it a second digital dividend [30]. This will be on the agenda for a final decision at WRC 2015. In the USA other moves have been afoot as the USA is attempting to clear more of the TV bands by using incentive auctions. A bill has been passed authorizing the Federal Communications Commission to hold incentive auctions for broadcast spectrum [13]. The bill "grants the FCC authority to conduct incentive auctions under which it shares some of the proceeds with licensees who return spectrum." The aim of the voluntary incentive auctions is to provide a means for current *licensees of valuable but underutilized frequencies* to be compensated for surrendering all or part of their spectrum, while making it available for high-demand mobile wireless services. In principle we could be facing a reduction in the amount of White Space that exists.

More challenging is the economies of scale issue. To reach an economy of scale that makes radios and other equipment for the TVWS worth manufacturing there has to be a pull from the market – i.e. a demand for the wireless services and applications that will use the infrastructure. But to deploy the infrastructure so that these services and applications can begin to materialise, the equipment needs to exist. This is the typical viscous circle that occurs in these situations. The fact remains that while a handful of companies are making equipment for this space, it currently remains hard to source at any scale. Standardised commercial equipment is still some way off – they are still at the early prototyping stage. Even though the IEEE has progressed both 802.22 and 802.11 af (WiFi for TVWS), commercially available products emanating from these standards are not yet ready. In addition any

---

[4]http://www.weightless.org/

equipment that is currently being manufactured is equipment that conforms to FCC or Ofcom rules rather than equipment that might potentially test the true value of the TVWS.

10.4 The Sharing Economy

The TV White Spaces are an opportunity that arose before the concept of sharing had matured. As such, the solutions proposed were somewhat idiosyncratic, ad hoc and done in reaction to pressure from various lobby groups. In spite of this, the process of regulating the TV White Space has shown that sharing *is possible* among a hierarchy of sharers. It suggests that the balancing of needs between primary incumbents (TV broadcasters), secondary incumbents (PMSE users) and general access cognitive radio users in a sharing paradigm is within grasp of spectrum managers.

While the sharing economy is being slowly adopted by the RF spectrum world, it has been embraced in the wider economy in recent years. There is an acknowledgement that high-value resources can not be left idle. The exclusive sharing, or collaborative consumption [1], of resources has been in existence for a long time, e.g. car rentals, hotel bedrooms. Fractional ownership is a concept that has been used to manage resources such as private aircraft or holiday villas. The basic business model of a hotel clearly illustrates the concept of collaborative consumption. This is an industry that invests in expensive real-estate, think Manhattan, builds a durable, luxurious, building containing hundreds of bedrooms and shares those beds among hundreds of thousands of users who otherwise most likely could not afford to either buy a property or take a short-term lease for the time they require to be in Manhattan. While hotels do not market their services as bed-sharing facilities, that is what they are.

The basic premise underlying these sharing models, whether it is for hotels (which share beds), car rental or aircraft sharing, is that an individual resource consumer cannot afford, or does not wish to invest in, the cost of developing and maintaining a resource for which it will not have a permanent continuous need. The reach of this model has spread in recent times; bike-sharing and car-sharing schemes have cropped up in many cities, private home/bed-sharing industry has grown that uses private homes to provide accommodation for strangers. In the case of the bike-sharing schemes as operated in Europe, municipal governments have provided advertising space in return for the investment in bike-sharing infrastructure, i.e. bikes, secure bike stands and payment/security mechanisms; city dwellers can then use the bikes for a nominal sum for a limited time. Such has been the success of this model that it is now being adopted in very large and wealthy cities such as New York. In the case of car-sharing schemes such as Zipcar and Streetcar users can access a car by the minute and by the hour, collecting and returning cars on streets convenient to them. The carsharing model is a more fine-grained car (resource) allocation model than the traditional car rental scheme in which rental is by the

day and organised from, often inconvenient, car lots. The success of this approach can be seen in the acquisition of Zipcar by Avis, an old-fashioned car rental firm. They see the need to provide a finer granularity of sharing which addresses more localised needs, alongside their traditional offering.

Sharing requires a mind-shift change from traditional network operators. It requires them to be comfortable with some degree of non-exclusive medium-term access to spectrum, alongside their traditional exclusive holdings, and with the impact that it *may* have on their business. It requires that traditional network operators need to become more confident in adaptive and cognitive technologies to deal with dynamic spectrum availability – even though the level of dynamism may be on a very slow basis, e.g. losing access to a portion of shared bandwidth for a day when given a week's notification.

So while there is still a demand for exclusively held spectrum in the telecommunications domain, borne out of familiarity as much as anything else, the supply of such exclusively-held spectrum will simply not be there. As such, spectrum users, network operators and spectrum managers need to embrace models of resource sharing that emerged in other domains. In the radio world, sharing and collaborative consumption will be enabled by advances in both radio and *non-radio* technologies, e.g. advances in database design, cloud management, geo-location, security and payment systems. These all serve to lower the transaction costs for new spectrum management frameworks that enable more complex sharing or trading arrangements.

10.4.1 International Moves: PCAST, LSA and BSOs

International regulatory action on spectrum sharing is not being progressed to excite researchers of cognitive radio. It is being done as there is little doubt that there is a looming wireless capacity *crunch*; as we noted earlier the forecasts vary wildly but they all indicate that current allocations for mobile wireless broadband are insufficient to meet the needs of 2020 and 2030. This capacity demand can only be met by driving efficiencies in all areas. That means increasing the spectral efficiency of existing networks by deploying better radios, more MIMO, etc. It means changing the network topology by making it more dense, i.e. relying on more small cell deployments and increasing the density of the maco-cell networks. And it means getting access to more spectrum. However, releasing more spectrum using the conventional method of exclusive licenses awarded through auction is arduous and time-consuming. And sharing, aided over time by maturing cognitive radio technologies, is seen as the only cost-effective way of releasing enough spectrum for mobile broadband use within an acceptable timeframe.

The US President, mindful of this *crunch*, commissioned a report, known as the President's Council of Advisors on Science and Technology (PCAST) report [29], to see how the government could take steps with its own spectrum to address the demand. The first PCAST conclusion was that clearing spectrum costs too much

and is not a sustainable way release spectrum for other, newer uses. The NTIA put the cost of clearing a band of just 95 MHz at 1,755 MHz at \$18 billion [21]. They also estimated it would take 10 years and cause significant disruption to the services being moved. The PCAST report indicates that even if the delays and disruption were acceptable, the annualised return to government from this approach would be so small that it could not support or motivate a continued policy of clearing and reallocation. The economic infeasibility of relying on this outdated spectrum management framework is a key driver to move forward with spectrum sharing and with cognitive radio.

In ditching the old and static spectrum management framework the PCAST report makes a series of bold recommendations to release 1,000 MHz of US Federal spectrum. These include the implementation of Spectrum Access Systems (SAS), basically enhanced versions of the TV White Space databases which would act as spectrum clearing houses providing rules and authorisations for different bands, new receiver requirements, test cities and a move to more tiered model of shared-access of Federal spectrum. There would be three tiers in this new spectrum management model; Federal Primary Access, Secondary Access and General Authorised Access (GAA). The Federal Primary Access users would register their *actual deployments* within a database. These Federal Primary Access users would enjoy full exclusive use of their assigned spectrum wherever they deployed their networks or systems but would relinquish rights to exclusivity away from deployments. Secondary Access users would be issued with short-term access rights and would also be registered in the database. These rights give them certain guarantees of levels of service and would assure them or protection from interference from the lowest users in the tiered access model. Within the Secondary Access tier it would be open for the regulator to create multiple levels of users which could have preferences over each other. The final and lowest layer of the hierarchy is the General Authorised Access tier. GAA users would be purely opportunistic, only being assigned whatever frequencies are left over at any given time and in any given location. If a Federal Primary Access user or a Secondary Access user were to request spectrum from the database the GAA user would have to vacate the bands. As such, the feasibility of GAA users being able to use such a system is very much predicated on them having access to frequency agile radios and cognitive radio capabilities.

The *aspirational* changes envisaged by PCAST cannot be implemented overnight, or even within a decade. Rather, they spell out a new way of managing spectrum which embraces cognitive radio and dynamic spectrum access technologies, be they in the limited forms that are currently found in new standards such as LTE or in the more advanced forms that are still very much at the R&D level.

One of these slow, but important, changes will arise from their recommendations to focus on receivers. Spectrum management frameworks have traditionally been transmitter focused, and they still are in the main. Licenses were awarded in terms of transmitter-related conditions for specific technologies and services. Interference management was dealt with using guardbands between different allocations and using frequency planning arrangements within assignments. Less crowded spectrum

assignments meant that legacy topologies employing fewer large transmitters could tolerate the use of cheap receivers in the user equipment. Standards bodies addressed minimum criteria for receivers to operate within distinct and stable spectrum allocations, e.g. TV receivers were designed to operate in the UHF bands without a care for other systems operating within or adjacent to those frequencies. As such, the effects of receivers on interference was largely ignored and not regulated.

This kind of laissez-faire attitude regarding receivers creates difficulties for spectrum management as sharing essentially removes much of the certainties associated with strict spectrum allocations. Sharing spectrum envisages allowing new and unforeseen radios and services to use spectrum assigned and allocated to uses and users that have subsequently turned out not to using their spectrum efficiently. A receiver that has operated adequately in a stable and non-shared situation may exhibit inadequate receiver performance if the transmitting radios on the adjacent bands change in such a manner that the mix of received signals overloads it. And even when systems are deployed along strict spectrum allocations and assignments unforeseen interactions can occur if the receivers are not of sufficient quality. The Nextel/Public Safety and Lightsquared/GPS problems in the US demonstrate the problems that can occur. Even though both the Nextel and Lightsquared networks were deployed in accordance with their license conditions, the quality of the receivers in the adjacent public safety and GPS networks were unable to cope. In the case of Lightsquared, they were forced to completely shut down their nascent network and have had their license suspended indefinitely.

To address these kinds of situations from occurring in more dynamic and shared spectrum environments, PCAST calls for improved receiver regulation. This means having better information about the characteristics of receivers operating in a given band and it means setting down minimum standards for receivers that will be deployed in future spectrum assignments. Firstly, if a spectrum management framework actually knows what the receiver performance characteristics of the incumbent radio systems operating in a band are, and the characteristics of receivers seeking to enter the band as sharers, then it can make better assignments which mitigate the possibility of negative adjacent band responses occurring. Secondly, by imposing receiver interference limits, i.e. setting down how much interference should be tolerated before it can be classified as *harmful* interference, something for which the regulator might have to find a remedy. The UK's Ofcom has already proposed a framework of Spectrum Usage Rights which would align very closely with this concept [26]. This focus on a more receiver-centric spectrum management framework will be quite dependent on cognitive radio developments. Frequency agile radios that can tailor their spectrum emissions in response to the needs of dynamically-assigned adjacent band/channel neighbours will be. As shared spectrum in turn becomes more congested, it will be the more nimble and accommodating cognitive radios that are able to avail of every last spectrum opportunity.

While the PCAST report pushes sharing as a possible solution it also acknowledges that certain types of innovation in terms of devices, standards and markets require different types (or grades in terms of quality) of spectrum for different

purposes. While many companies thrive in the unlicensed spectrum spaces, others that may seek to challenge the incumbents who hold exclusive licenses have fewer opportunities to access similar exclusive spectrum rights so that they can attempt to innovate their services. Basically, exclusive spectrum rights, especially any bands earmarked for IMT/IMT-Advanced services, are grabbed by the major mobile network operators leaving virtually nothing for start-ups. The PCAST report questions how a venture-backed start-up could be supported in attaining 6 or 12-month access to spectrum and prove itself to the market, possibly proving the use of new cognitive radio technologies that the dominant standards are not readily embracing. Essentially, the report suggests that while there is demand at both ends of the market, for long-term licensed spectrum and for instantaneous unlicensed spectrum, there is also likely demand for spectrum somewhere in the middle of that continuum, i.e. for short, medium and long-term licences.

But the FCC, Ofcom, the EU have already put in place regulations to allow for the secondary trading of spectrum. Secondary spectrum trading is, theoretically, a means by which holders of tradeable exclusive licenses could trade part or all of their spectrum rights with other parties. The regulations generally allow for the primary licences to be partitioned such that the primary licensee can release excess spectrum to a lessee for a specified time. This approach gives legal certainty and guarantees to both parties, the incumbent and the innovative start-up. However, the take-up of trading has been very slow. There are number of reasons for this. Firstly, there is a lack of open network infrastructure, e.g. wholesale RAN infrastructure, on which new entrants can use speculative or short-term spectrum. Secondly, trading of spectrum in this manner has the potential to fragment spectrum bands among operators which is contrary to how operators and regulators currently plan its use. Currently, there is a lack of technologies that can readily embrace fragmented spectrum opportunities though this is changing with the carrier aggregation capabilities of LTE-A. Finally, commercial license holders may not want to admit that they are not fully utilising their spectrum, for reasons of commercial sensitivity, and they may also be incentivised to block the innovation of competitors by denying them access to spectrum.

The PCAST report acknowledges this and sees that there may be a role for government holders of spectrum to play by making Federal spectrum accessible in accordance with new spectrum management principles. It suggests that there may be a place for three types of licenses in a new spectrum management framework for Federal spectrum. *Long-term licenses* would be very much along the lines of the current FCC-issued Personal Cellular Service (PCS)/Advanced Wireless Services (AWS) licenses with 10–15 year durations, but with clauses that allow for well-defined easements for low-power unlicensed use or with pre-emption conditions for public safety networks. The later idea closely aligns with Nancy Jesuale's suggestion for infrastructure and spectrum sharing in the D-block which leverages DSA technologies [18]. *Medium-term licenses* would be for terms of 3 years or less and would be issued on a negotiated case-by-case basis to allow for new, bespoke uses of spectrum so that ideas, innovations can be tried. Notably, there would no automatic renewal of these licenses. This approach could allow for a

better turnover of spectrum among commercial users. The final category is *short-term licensing*. These licenses would essentially cover pre-packaged spectrum rights with fixed, non-negotiable terms and conditions lowering the transaction costs. Different packages could have different associated quality of service and could be available with different temporal, frequency and geographic dimensions. This kind of spectrum could be sold through enhanced databases and its use is quite dependent on the development of frequency agile, cognitive radio-enabled networks that can exploit short-term spectrum licences that vary by both frequency and quality.

The key message that PCAST sends to policy-makers is that the simple model of accruing revenue from license fees by selling cleared spectrum, or indeed selling short and medium-term licenses, should not be the objective of adopting a new framework. Rather, the framework's objective should be to enable maximal spectrum efficiency through supporting a mix of spectrum access/rights regimes which support different types of technologies and business models at differing points of maturity. To return to the language used in the outset of the chapter, the PCAST report covers a wide variety of ways in which spectrum can be consumed. If policies were put in place, it would lead greater exploitation of the excludable/non-excludable and rival/non-rival continuum and therefore a richer array of options for accessing spectrum.

Increased demands for capacity are also looming in Europe. The European bodies responsible for drafting and adopting new radio regulations, the Commission, CEPT, the Radio Spectrum Policy Group, etc., are also trying to formulate and formalise a framework that will allow for more sharing of spectrum.

Europe has been looking at increasing access to sharing through two approaches; the more typical unlicensed sharing approach akin to the ISM band sharing where an unlimited number of users can access a band once they operate to some minimal non-negotiable technical requirements, and another concept built around licensed sharing where a limited number of individual users are allowed to share a band of spectrum according to a set of negotiated rules and conditions. The later approach to sharing spectrum is known as Licensed Shared Access (LSA); it was initially proposed by an industry consortium under the name Authorised Shared Access (ASA). This framework would see the initial licensed user or users, i.e. the *incumbents*, sharing *their* spectrum with one or more new users who may be offering the same service or a different one, according to conditions imposed on both the incumbent and new user.

The EU's approved Radio Spectrum Policy Programme endorsed the LSA concept and CEPT is fleshing out the details of how such a system would work in practice, both within the confines of technology and with the confines of the European Treaties [3]. LSA would introduce a number of new actors/entities to the regulatory landscape. In addition to the NRA and the incumbent there would also be potential *LSA Licensees*, an *LSA repository*, and an *LSA Controller* [2]. In the model being mooted by CEPT, the NRA would maintain a traditional role of awarding LSA Licenses to a number of operators who want to share the spectrum band of an incumbent. How many LSA Licenses a NRA awards, and manner in which it awards them, i.e. whether by auction or beauty contest, is not pertinent to the operation of

the scheme. Those operators that receive LSA Licenses become LSA Licensees. The details of the License will indicate how the incumbent and the Licensee will share the spectrum; this could involve details of adjacent channel protections, geographic exclusion zones, etc. These details could be worked out bilaterally between Licensee and the incumbent or they could be imposed by the NRA.

The LSA Repository is a key component in this system. It is basically akin to the TV White Space database, describing the available spectrum, by frequency and geographic area, where LSA Licensees could potentially operate. As with the TV White Space database the LSA Repository is not necessarily an honest or true reflection of the spectrum available. Depending the interference protection criteria chosen and modelled for different LSA Licensee/Incumbent interactions less spectrum than could actually be used may be revealed in the repository. Furthermore, to guard against commercial and competition concerns it may not be prudent to reveal the incumbent's actual use of spectrum by reflecting it honestly in the database. Nonetheless, the LSA Repository is the key component describing what actual spectrum resources may be shared at a given time.

The LSA Controller operates on behalf of a LSA Licensee. It translates the information provides by the LSA Repository as to what spectrum is available where and for how long, and sends that information to those elements of the LSA Licensee's Radio Access Network that will configure the radio systems to use the licensed shared spectrum. There are *many, many* details to be worked out and in Sect. 10.4.3 we describe some work that has been progressed for LSA in the 2.3 GHz band.

However, as we noted earlier, incumbent operators have been slow to trade their unused spectrum using the secondary trading regulatory frameworks that are available to them under EU law [6]. Indeed, many NRAs have been slow to adopt and implement supporting legislation and mechanisms in the EU's Member States. Just as operators are unwilling to trade their spectrum, they are also likely to be unwilling to share it. Again, a willingness to share may indicate that a firm's spectrum holding is not fully and efficiently utilised across all frequencies, at all times and in all places. It may also give a competitor an edge. As such, if regulators are to rely solely on voluntary sharing arrangements to emerge and satisfy the projected demands they may not meet their targets. To this end the European Commission have initiated process to create a spectrum management framework which will identify potential sharing opportunities such that NRAs could force spectrum sharing where it is economically and technically feasible.

The European Commission have set a simple criterion to establish if sharing is an economically viable option for a given spectrum band. They consider that a Beneficial Sharing Opportunity (BSO) exists wherever the combined net socio-economic benefit of multiple applications (including any incumbent) sharing a band is *greater* than the net socio-economic benefit of the single incumbent, taking into account the costs of facilitating such shared use. Making such a judgement requires knowledge about the kind of sharing arrangement that would be put in place; would there be a sharing hierarchy between incumbents and newcomers or would they be equal, what standards (e.g. resilient receivers) or protocols would be mandated.

To allow for economies of scale within the EU itself the process of identifying BSOs should be standardised across all the Member States. Regardless of whether it happens in harmonised or non-harmonised bands the Commission feel that a certain level of coordination is necessary to support sharing on such a scale.

However, the progress of the BSO and LSA concept will require that a number steps are taken. In advance of even contemplating whether the cost of setting up a LSA scheme is justified for a given band, the NRA or regional authorities should have some indication as to the amount of spectrum activity and efficiency that is associated with that band. But existing network operators do not, as a rule, publish information about their own spectrum activities and nor are they required to. NRAs themselves do not usually conduct an efficiency inventory of every band at every location within their jurisdiction. Knowing who is using spectrum, where they are using it and how well they are using it is a key building block in pursuing the identification of BSOs. Without access to this knowledge it will not be possible to cost sharing frameworks suited to different bands, to understand the implications of sharing on existing spectrum users and to forecast the effects of innovation, sharing and increased competition on any band. In the absence of access to open spectrum inventory across the EU Microsoft has commenced an initiative to observe, collect, collate and disseminate spectrum usage data through its Spectrum Observatory project.

This very slow approach to analysing spectrum usage through spectrum observatories, identifying BSOs through transparent processes, allowing for sharing through some type of LSA is a form of Dynamic Spectrum Access in *slow-motion*. However, it introduces and formalises the necessary concepts to the regulatory and commercial spheres and portends of a *faster* DSA environment enabled by cognitive radio technology. Once the kinks have been removed from the initial processes there is little reason that the spectrum observatory is not replaced by CR radios in the field identifying BSOs and entering in LSA agreements over the course of days rather than months and years.

While both the PCAST report and the European embrace of LSA are at a very high level, quite removed from implementation in actual regulation and adoption by industry there are at least two current instances where policy is being worked out in some detail to address the aspirations outlined in this section. These are the moves to regulate sharing in the 3.5 GHz band in the USA and moves to regulate LSA usage of the 2.3 GHz band in Europe. In the next two section we describe the ongoing work.

10.4.2   The 3.5 GHz Sharing Opportunity

One of the PCAST recommendations which is having tangible effects concerns opening of the 3.5 GHz band (3,550–3,650 MHz) in the US for shared access. An NTIA report had identified this band as suitable to be fast-tracked for sharing between US Federal and non-Federal users and uses [25].

The FCC are now leading the regulatory consultation process for this band, turning the aspiration of the PCAST report into law. An FCC notice of proposed rule-making (NPRM) [14] proposes the creation of a new Citizens Broadband Service (CBS) in the 3.5 GHz Band. The objective of this consultation process is to open up the band to low-power small-cell users. As this is a shared band the regulations must take account of the incumbent systems. Neither the in-band nor adjacent band incumbent systems were designed or deployed with this sharing in mind. In the lower band there are incumbent high-power ground and airborne military radar. Within the band there are high power Federal Department of Defence radars located along the Eastern, Western and Southern coasts of the Continental USA. There are also non-Federal users; there are 37 Fixed Satellite Service (FSS) earth stations (receive only mode) located in various cities through the country. Any new users of this band cannot degrade the performance of any of these services. Originally, the NTIA had done some modelling around sharing this spectrum which used WiMAX as the candidate technology (WiMAX was commercially available for this band at the time). Using a macro-cell type topology at this frequency they found that exclusion zones covering approximatively 60 % or the population would have to be enforced to protect the incumbent services. This would obviously diminish the interest of commercial parties; device manufacturers, network operators, etc. However, 3.5 GHz band is ideally suited to small cell radios and topologies. The line of sight (LOS) properties of 3.5 GHz mean that signals from radios placed indoors will be highly attenuated and localised. In effect it means that the areas covered by exclusion zones could be drastically reduced. This again highlights the importance of the choosing the right numbers and right techniques to populate any database or repository that is used to indicate what spectrum is free or occupied.

The three proposed tiers of the FCC's 3.5 GHz CBS system are Incumbent Access (IA) which equates to PCAST's Federal Primary Access, Priority Access (PA) which equates to PCAST's Secondary Access, and General Authorized Access. Under the FCC's proposal, users in the PA and GAA tiers would be licensed-by-rule as Citizens Broadband Service users under Part 95 of the Commission's rules. It is intended that some form of Spectrum Access Service (SAS) be created to govern the interaction between the three tiers and this SAS would incorporate a geo-location enabled dynamic database and, potentially other appropriate mitigation techniques. The FCC have suggested that leveraging the knowledge and experience of the TV White Space database developments could accelerate the process of regulating the 3.5 GHz band.

The Incumbent Access tier would consist solely of authorized federal and grandfather-licensed FSS 3.5 GHz Band users and would have protection from interference from all users. The PA tier would consist of a portion of the 3.5 GHz Band designated for small cell use by certain critical, quality-of-service dependent users at specific, targeted locations. PA users would be eligible to use authorized devices on an interference-protected basis as determined by the SAS. The PA tier would be available only in areas where the PA systems would not cause interference to incumbent operations and would not be expected to receive interference from incumbents. These are referred to as PA Zones in the FCC NPRM. In the GAA

tier, licensees would be authorized to use the 3.5 GHz Band on an opportunistic basis within designated geographic areas. Use of GAA devices would be permitted in PA Zones as well as areas where such devices would not cause harmful interference to incumbent operations but where signals from incumbent operations could be expected to interfere with GAA uses on occasion. There is a clear sense from regulators that enticing operators, who would historically have preferred exclusively-licensed spectrum, into the shared spectrum domain means that they have to ensure that they can deliver some guarantees of levels of service. Otherwise, there is every chance that they would end up with a two-tier system; Federal Primary Access users and General Authorisation Access users.

10.4.3 The 2.3 GHz Sharing Opportunity

As we noted earlier, Licensed Shared Access (LSA) is a concept being developed and matured by CEPT for use in Europe [32]. CEPT now has two working groups focusing on bringing LSA rules to fruition. CEPT's Electronic Communications Committee Working Group (Frequency Management) 52 is working to develop harmonised rules for the 2.3 GHz band; that includes the band plan, the specific rules for LSA within the context of the band. CEPT's WG 53 is working in a more general sense to advance understanding and rules around TV White Spaces, LSA and Reconfigurable Radio Systems. It also interacts with ETSI to further these aims.

The reason that there is a drive to formalise rules for LSA in the 2.3 GHz band is driven by the demand of traditional mobile network operators (MNOs) who seek more and more harmonised spectrum for LTE/-A. Releasing spectrum on an exclusive basis to mobile operators is still the preferred option of European regulators. It is currently the simplest and cheapest arrangement in terms spectrum management. However, given that the demand for IMT-Advanced spectrum allocations is exceeding the supply of globally harmonised bands the regulatory authorities will be unable to meet the needs of LTE in a timely and cost-efficient manner by relying on clearing and reallocation. The 2.3 GHz band (2,300–2,400 MHz) has already been identified by the ITU Radiocommunication Sector for IMT use and LTE-Time-Division Duplex systems are already deployed in this band in Asian and Middle Eastern countries. This means that equipment already exists and European operators would benefit from the economies of scale were they able to access this band as well.

However, the use of the 2.3 GHz band is not harmonised at a European level. Different Member States have made various uses of the band. In some countries, such as Ireland, it is a cleared band and so it can be assigned to operators using the conventional market-based exclusive assignment methods. In other countries it has incumbent users such as defence systems, wireless camera operators, etc. As the process of harmonising the band would be too slow and too costly the aim is to develop LSA regulations for those countries which have incumbent users. In time, these incumbent users may migrate to other bands as the natural life of their

services/technologies ends and other bands are made available to them. This can be thought of as a slow form of clearing and re-allocation or *soft harmonisation*.

CEPT is focused on developing a band plan and LSA rules that would allow Mobile/Fixed Communication Networks (MFCN) users to access the band. In effect these will likely be LTE operators and this is the candidate technology that CEPT is using to develop regulations for the band. Currently, the proposal sees the band plan divided into 5 MHz channels on which will be imposed the Least Restrictive Technical Conditions (LRTC). This is an approach which has been mandated since the adoption of the Wireless Access Policy for Electronic Communications Services (WAPECs) directive [27]; it seeks to have the most technology and service neutral conditions imposed on licenses where possible so that the market can decide on the best use of the spectrum rather than the regulator picking the winner, and possibly failing. In line with recent decisions the Block Edge Mask (BEM) will likely be the chosen method to control out-of-band/block emissions. A BEM is a an emission mask that is defined as a function of frequency relative to the edge of the block of spectrum that is licensed to an operator. This technology and service neutral approach gives the operator some freedoms to decide how it plans its network; cognitive radios are particularly apt to exploit BEMs [15, 19]. While such a band plan may be optimal way to coordinate new users in a cleared band, it does not address the incumbent users who not harmonised to such a plan. So while a new operator may be assigned a number of 5 MHz channels within the band, the right to use this channels might be abridged from time to time under the sharing model of the LSA system.

Consequently, one of the key attributes, and initial worries, of an LSA system will be its ability to deliver on guarantees of QoS or 'levels of service'. As spectrum rights given under LSA spectrum differ from the rights preferred by commercial operators such as LTE MNOs, regulators are keen to ensure that they create an attractive LSA system. To that end, the management of interference from LSA band incumbents *to* LSA sharers is just as much of a concern, if not more, than interference/interruption from the new LSA licensees into the incumbents. There is a sense that CEPT do not want potential high-value commercial customers of LSA spectrum to be mis-sold spectrum access and then put off the concept of shared spectrum rights.

In focussing on the 2.3 GHz band as a starting point for developing a detailed LSA architecture and rules, the CEPT WG have to address a range of specific challenges that the abstract LSA concept does not reveal. The first challenge concerns interference to and from the various other systems that sit adjacent to the 2.3 GHz and within the band. As noted earlier, LSA, and sharing in general, has the potential to introduce radio systems which neighbouring systems did not anticipate. The 2.3 GHz band has a range of adjacent neighbours; below it there are Deep Space research uses including space-to-earth and space-to-space applications, radio astronomy and defence applications; above it lies the ISM band. Dealing with the adjacent neighbours should be relatively straightforward; CEPT can adopt the BEMs at the 2.3 GHz edges that control interference out of the band, limiting a negative adjacent band response from incumbent systems. The CEPT WG can also

decide on geographic exclusion zones around earth stations in the adjacent bands which, further limiting the risk of inference. But as noted in the PCAST report, sharing could improved if better information was available about the actual receiver characteristics of potential victim systems in the adjacent bands or, indeed, in-band and also if those receivers had actual receiver interference limits which indicted what harmful interference was. In would allow the LSA system to dynamically adjust the condition it imposes on sharers of the band such that it maximises their efficient use of the LSA spectrum whilst avoiding any harmful inference.

It is the introduction of sharing *within* this band presents the more complex and more interesting picture for LSA. In addition to fixed services such as Aeronautical Telemetry the 2.3 GHz band is also used for Services Ancillary to Programme-Making/Services Ancillary to Broadcasting (SAP/SAB). These services are generally the wireless cameras used for outside broadcasts at sports events, crime scenes, etc. and also those used for film-making, concerts, etc. While these services also use frequencies outside the 2.3 GHz band they will continue to be treated as incumbents of the 2.3 GHz under the LSA regime. The nature of SAP/SAB use can be quite dynamic. In addition to being highly localised to an event, the timing of use can vary from *very short notice*, such as for coverage of an emergency situation, to *scheduled*, such as for a special event like a parade. While localised in terms of geography, the area covered can vary from a small area, e.g. outside a courthouse, to a large area, e.g. swathes of a city during a marathon.

As with the adjacent band neighbours outside the 2.3 GHz band, the CEPT WG is determining *both* the geographic exclusion zones that would be required to protect SAP/SAB users from *co-channel interference* from a LSA sharer, which might be up to 33 km for portable video links, and also the *adjacent channel interference* protections. What we can say about these exclusions is that they will have different dynamic properties; some may be scheduled in advance and some may be abrupt. They will be localised, relative to a country-wide jurisdiction, but they may range from city-block to an entire town. And the frequencies that need to be released may not align neatly with the harmonised band plan imposed on the new LSA entrants.

Since the new entrants are likely to be LTE operators the kinds of exclusions likely to protect occasional SAP/SAB users will pose new challenges to LTE operators. LSA is being progress by CEPT on the basis of two simple assumptions. Firstly, that LSA user is always using the released spectrum when it is not required by the incumbent. Secondly, that authorised LSA operators use of the band is as widespread as possible. These two assumptions pose two questions; how quickly can the LSA operator be notified that the incumbent (e.g. the SAP/SAB) needs to access the spectrum and switch off its services; and over what area should the LSA operator need to switch off its network. In practice, for the case of SAB/SAP, spectrum availability can be withdrawn from the LSA Licensee by amending the LSA Repository to take account of the need the incumbent user's needs. The CEPT WG anticipates that the 2.3 GHz LSA Controller would then signal to the operations, administration and management of the relevant LTE networks. The exact modalities of this process are still being explored.

The notion that a user of a wireless camera at a scene of a road-traffic accident would have the right to have a MNO's macro-cell switched off would likely have put many network operators off the concept of LSA. However, LTE architectures are inherently designed to support many new types of topologies. Small cell topologies are already a growing feature of many networks. Such topology, if it is using a mix of pico- and femto-cells, would be much more resilient to an LSA request for a localised exclusion. Small cells employing sophisticated MIMO should be able to adapt the network's coverage at the relevant frequencies to avoid the LSA exclusion area. Furthermore, the advent of HetNet and Carrier Aggregation in LTE means that operators will have the technology to deal with variable frequency availability. The HetNet approach will allow an operator to seamlessly knit together a small cell network topology operating in one band with a macro-cell topology operating in another band. As such the LTE HetNet approach allows operators to mix their use of consistent, uninterruptible exclusively-held spectrum with possibly interruptible LSA spectrum such that they can balance the loading on their network between the two in a dynamic manner.

10.4.4 The Outlook

Overall there seems to be movement in different pockets internationally which support some new forms of sharing. Returning to the language introduced at the outset of the chapter, the rules and policies which are under consideration for accessing spectrum increasingly embrace some aspects of sharing and in combination with new technologies can exploit other opportunities within the excludable/non-excludable and rival/non-rival continuum, i.e. around the greyed out area in Fig. 10.2. There tiered forms of access indicated in PCAST and the possibilities within ASA/LSA can be varied and rich. Trials are already underway for the latter for example. The Finnish CORE+ (Cognitive Radio Trial Environment +) project consortium has demonstrated the world's first trial of the ASA concept for spectrum sharing, using live an LTE/4G trial environment operating in Ylivieska, Finland. Hence the progress from a technical perspective is rapid. There is a danger however, that sharing of spectrum simply becomes a means of allowing the mobile operator incumbents ONLY, gain access to more spectrum and that these new and interesting modes of accessing spectrum are not open to a wider range of players.

10.5 Other Work of the IEEE

This discussion so far has broadly focused on the TV white spaces and on the emergence of spectrum sharing initiatives in different regulatory jurisdictions, mainly driven by Government authorised bodies. For the sake of giving a complete view it is worth noting some other activities. As mentioned in Sect. 10.2.1 the IEEE-SA

is not a formally authorised government body and tends to work through a more global and community oriented approach of gathering interested parties together to work on topics. The IEEE 802.22 standard was in fact included in the discussion on TV White Space. ThIEEE-SA is active in a wider range of activities pertaining to cognitive radio and dynamic spectrum access. The IEEE DySPAN Standards Committee (DySPAN-SC) provides the focal point for a large amount of this work.

10.5.1   IEEE DySPAN-SC

The Dynamic Spectrum Access Networks (DySPAN) standards committee, formerly Standards Coordinating Committee 41 (SCC41), and more formerly still the IEEE P1900 Standards Committee, is sponsored by the IEEE. Its working groups and resulting standards, are numbered in the 1900 range and can be collective referred to as IEEE 1900.X. More specific details of the name changes and the bodies in control can be found of the IEEE-SA website.[5] There are seven different flavours of P1900 working group, each devoted to a different set of topics.

1. IEEE P1900.1 This working group focuses on the definitions of and the concepts surrounding dynamic spectrum access. While a document of terminology was published in 2008, extensions to the terminology and amendments are being made under the heading 1900.1a.
2. IEEE P1900.2 This working group focuses on in-band and adjacent band coexistence issues and interference issues. The IEEE P1900.2 standard has been published and provides technical guidelines for analyzing the potential for coexistence or interference between radio systems operating in the same frequency band or between different frequency bands.
3. IEEE P1900.3 The working group was focused on conformance evaluation of software defined radio modules. However it has been disbanded.
4. IEEE P1900.4 The working group was a very active working group and focused on architectural building blocks for network-device distributed decision-making for optimized radio resource usage in heterogeneous wireless access networks and a standard has been published. Current projects fall under the heading of 1900.4a and 1990.4.1. The former is making amendments to the initial standard to include architecture and interfaces for dynamic spectrum access networks in white space frequency bands. The latter is looking at a standard for interfaces and protocols which enable distributed decision-making for optimized radio resource usage in heterogeneous wireless networks.
5. IEEE P1900.5 The purpose of this standard is to define a policy language and associated architecture requirements for interoperable, vendor-independent control of dynamic spectrum access functionality and behaviour in radio systems

---

[5]http://grouper.ieee.org/groups/dyspan/index.html

and wireless networks. This standard will also define the relationship of policy language and architecture to the needs of at least the following constituencies: the regulator, the operator, the user, and the network equipment manufacturer. The P1900.5 standard has been published but there is follow on work under the heading of P1900.5a which embraces amendments as well as projects entitled P100.5.1 and P1900.5.2. P1900.5.1 focuses specifically on the vendor independent policy language. The latter focuses on developing a common method for modelling spectrum consumption.

6. IEEE P1900.6 This working group created a standard for spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems. Like the other groups extensions and advancements of the standard are coming from the P1900.6a. It adds specifications for the exchange of sensing related and other relevant data and specifies related interfaces between the data archive and other data sources.

7. IEEE P1900.7 This working group is currently developing a draft standard for the radio interface for white space dynamic spectrum access radio systems supporting fixed and mobile operation.

As is evident from the brief descriptions given above there are a number of topics which bleed into the issues that have already been discussed in this chapter. For example issues relating to co-existence and adjacent and co-channel interference are very relevant for spectrum sharing. It remains to be seen which standards if any will catch on. A point which is made consistently through-out the chapter is the fact that cognitive radio and spectrum sharing is still very much an emerging area and much has still to be transferred to actual practice.

10.6 Some Perspectives for the Future

To finish off the chapter there are a number of perspectives worth highlighting which can perhaps provide additional thoughts on future opportunities.

10.6.1 Systematising of Sharing and Cognitive Technologies

As mentioned at the outset of this chapter a cognitive radio can simply be seen as an advanced form of radio requiring no specific regulations as 'cognition' becomes a natural part of the roadmap for future mobile communication systems. The example of LTE was furnished here. As could be any amount of self-organising networks. However rather than rely on natural evolution of these systems to embrace cognitive capabilities, it is also possible to proactively include features in future iterations of the standards that more explicitly target cognitive radio and sharing.

As an example of what this means consider carrier aggregation. Carrier aggregation, in very simple terms, allows blocks of spectrum to be aggregated to make larger
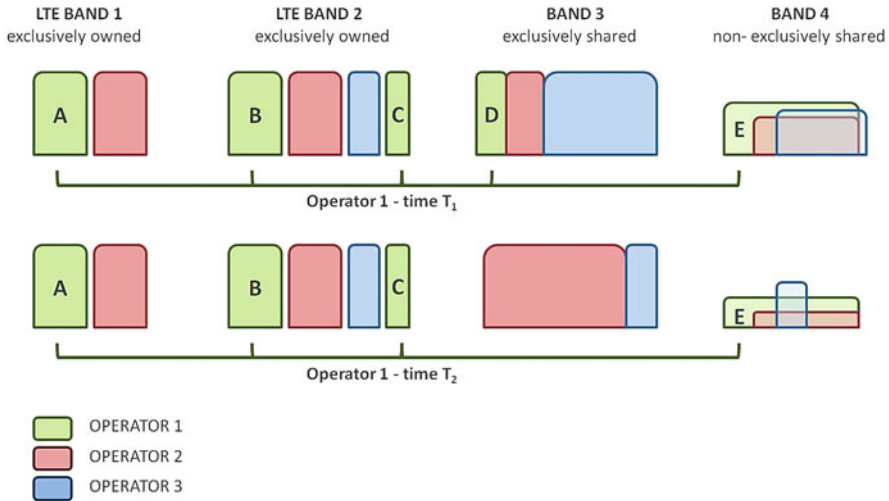
Fig. 10.3 An enhanced view of carrier aggregation

bandwidths available for the communication process. LTE-Advanced caters for aggregation of different blocks of spectrum within LTE bands or between different LTE bands. A chunk or block of frequency in the LTE carrier aggregation world is referred to as a component carrier. Carrier aggregation can be performed using contiguous or non-contiguous component carriers. Furthermore the component carriers may come from the same LTE band or come from different LTE bands. The former is referred to as *intra-band carrier aggregation* while the latter is known as *inter-band carrier aggregation*. Carrier aggregation, as currently defined, is envisaged as the aggregation of blocks from spectrum resources which are *statically assigned to a mobile network operator*. It is however possible to consider carrier aggregation in a broader context than is described in the LTE standards. There is huge potential for considering dynamic assignments, shared assignments and assignments that involve a mix of licensed and unlicensed access modes. Hence in [7] we introduced an advanced form of carrier aggregation that caters for scenarios in which carrier aggregation involves combining blocks of spectrum which are *spectrum blocks that are statically and/or dynamically exclusively assigned and/or blocks that are non-exclusively shared*. Figure 10.3 depicts this wider view of carrier aggregation. In Fig. 10.3 three mobile operators are depicted. The three operators have static assignments from two different LTE bands – which they can of course aggregate as they please. There are also now two additional bands available and these can be added to the aggregation mix. They also avail of exclusive time-based dynamic assignments as well as assignments from a commonly shared pool. Based on the ideas introduced earlier in the chapter it is not impossible to imagine the shared chunks of exclusively shared spectrum coming from pooled spectrum sources of spectrum or via some licensed shared access scheme while the non-exclusively shared coming from TV White Space bands.

❶
strong incumbent
= lone primary
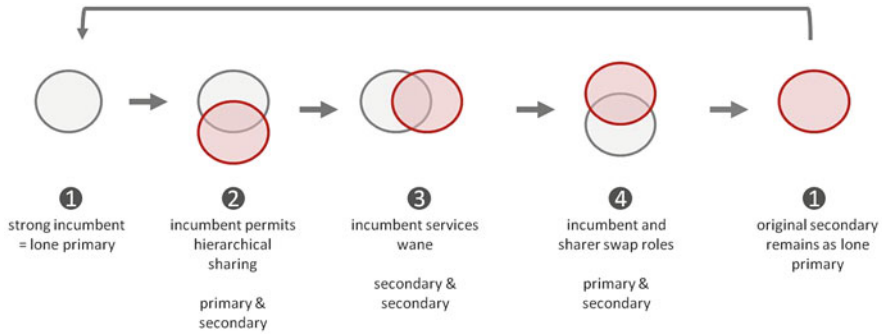
Fig. 10.4 Sharing is clearing

The purpose of introducing this example is not to argue specifically in favour of this advanced form of carrier aggregation but to show that it is possible to build cognitive radio concepts more aggressively into the LTE framework or any standard that is born of more traditional exclusive approaches. This system suggested in Fig. 10.3 also introduces a mix of rules that again make more use of the various options available to manage spectrum and support sharing, i.e. exploit other opportunities within the excludable/non-excludable and rival/non-rival continuum, i.e. around the greyed out area in Fig. 10.2.

10.6.2   Sharing Is Clearing

The second perspective that is worth emphasising relates to both sharing and clearing. Typically sharing and clearing of bands are set up as opposite view points. Those who wish to have complete control over bands prefer the clearing option. Clearing refers to the moving of incumbents out of specific bands and in most cases this is done to allow the empty bands to be auctioned for exclusive usage. The Digital Dividend for example refers to the band of spectrum that was cleared when TV services in the upper TV bands were re-located to lower bands. Clearing can be a costly and slow process. In some cases incumbents must be paid to move to other bands. And typically nothing happens in a band until the incumbents have moved. The times involved can vary from years to decades.

Sharing can be seen as a dynamic form of clearing. Figure 10.4 explains how. Initially a strong vibrant service is in operation in a band. As the need for this service wanes or because of the existence of large amounts of idle capacity within the band, sharing can be permitted (on a hierarchical basis for example). Over time it may be the case that the incumbent needs increasingly less spectrum – either as mentioned already because the needs for the service reduces or as for example in the case of Digital TV, more effective use of the spectrum is made. Roles may

then be reversed. Using the terminology of primary and secondary user, the primary incumbent starts as primary and exclusive user, subsequently permits sharing with a secondary system on some kind of hierarchical basis, over time both become equal sharers of the band and finally the secondary becomes the new primary as its services gain popularity and the old primary assumes secondary status as its popularity decreases until it finally dies out. Taking this approach means that there is a constant recycling of bands.

It could be argued that the suggestions for future use of the 2.3 GHz band in Europe and the 3.5 GHz in the USA are all about sharing until the incumbents completely release the bands.

10.7   Conclusion

This chapter provided a regulatory perspective in cognitive radio through focusing on the polices and rules that exist and are emerging around spectrum sharing. We consider cognitive radio as a key enabling technology for sharing and hence sharing regulations to be the most pertinent focus of the chapter. Overall the situation is challenging. The supposedly 'low hanging fruit' of the TV White Spaces has not turned out to be so low hanging. The pace of progress is slow and limited jurisdictions have adopted specific rules and regulations. However there are test and trials ongoing internationally and groups advocating for progress in this space. There are promising suggestions for moving forward emerging and potential for action in the 2.3 and 3.5 GHz bands in the medium term. Overall though as highlighted in Fig. 10.2 we are continuing to consume spectrum in limited ways rather than truly exploit the full range of opportunities that might not alone make the most of a scarce resource but also open the spectrum to new and more players. It is therefore hugely important that white spaces continue to be identified so that opportunities for exploitation of these white spaces can be amassed and the momentum for action increased.

There is of course also an inescapable reality that the cognitive capabilities will be a necessary part in some form or other of future communication systems. Hence whether the true power of the cognitive radio in unleashing dynamic and multi forms of shared spectrum access is delivered there is still a future for cognitive radio technologies themselves.

References

1. Botsman R., Rogers R.: What's mine is yours: the rise of collaborative consumption. Harper Business, New York
2. CEPT ECC Project Team FM53: Generic description of licensed shared access (LSA). FM53(13)11, Jan 2013

3. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Promoting the shared use of radio spectrum resources in the internal market, Sept 2012

4. Definitions of software defined radio (SDR) and cognitive radio system (CRS). Radiocommunication Sector of the International Telecommunications Union. ITU-R SM.2152 (2009)

5. Digital Dividend: Cognitive access. Statement on licence-exempting cognitive devices using interleaved spectrum. The Office of Communications (2009)

6. Directive 2009/140/EC of the European Parliament and of the Council of 25 Nov 2009

7. Doyle, L., McMenamy, J., Forde, T.K.: Regulating for carrier aggregation & getting spectrum management right for the longer term. In: 2012 IEEE International Symposium on Dynamic Spectrum Access, Seattle, p. 10 (2012)

8. 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

9. ECC Report 185: Complementary report to ECC report 159 further definition of technical and operational requirements for the operation of white space devices in the band 470–790 MHz, Jan 2013

10. ECC Report 186: Technical and operational requirements for the operation of white space devices under geo-location approach, Jan 2013

11. FCC: Second memorandum opinion and order: unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band. Federal Communications Commission. FCC 10-174, ET Docket No. 04-186 and 02-380 (2010)

12. FCC: Third memorandum opinion and order: unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band. Federal Communications Commission. FCC 12-36, ET Docket No. 04-186 and 02-380 (2012)

13. FCC: NPRM: expanding the economic and innovation opportunities of spectrum through incentive auctions. FCC 12-118. ET Docket No. 12-268 (2012)

14. FCC, NPRM: Amendment of the Commission's Rules with regard to commercial operations in the 3550–3650 MHz band. FCC GN Docket No. 12-354 (2012)

15. Forde, T.K., Doyle, L.E., Ozgul, B.: Dynamic block-edge masks (BEMs) for dynamic spectrum emission masks (SEMs). In: 2010 IEEE Symposium on New Frontiers in Dynamic Spectrum, Singapore, pp. 1, 10, 6–9 Apr 2010

16. Freyens, B.: The economics of spectrum management: a review' paper commissioned by the Australian Communication and Media Authority (ACMA), June 2007, p. 40. http://www.acma.gov.au/WEB/STANDARD/pc=PC_311025

17. Hardin, G.: The tragedy of the commons. Sci. New Ser. **162**(3859), 1243–1248 (1968)

18. Jesuale, N.: Lights and sirens broadband – how spectrum pooling, cognitive radio, and dynamic prioritization modeling can empower emergency communications, restore sanity and save billions. In: 2011 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Aachecn, pp. 467,475, 3–6 May 2011

19. Macaluso, I., Ozgul, B., Forde, T., Sutton, P., Doyle, L.: Spectrum and energy efficient block edge mask-compliant waveforms for dynamic environments. IEEE J. Sel. Areas Commun. **PP**(99), 1, 15, 0

20. Mittola J., Maguire G.Q.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**, 13–18 (1999)

21. National Telecommunications and Information Administration: An assessment of the viability of accommodating wireless broadband in the 1755–1780 MHz band. Washington, D.C., Mar 2012

22. Notice Of Inquiry: Additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band. Federal Communications Commission. FCC 02-328, ET Docket No. 02-380, Dec 2002

23. NPRM: Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies. Federal Communications Commission. FCC 03-322, ET Docket No. 03-108 (2003)

24. NOI: Promoting more efficient use of spectrum through dynamic spectrum use technologies. Federal Communications Commission. FCC 10-198, ET Docket No. 10-237 (2010)
25. NTIA: An assessment of the near-term viability of accommodating wireless broadband systems in the 1675–1710 MHz, 1755–1780 MHz, 3500–3650 MHz, and 4200–4220 MHz, 4380–4400 MHz bands (President's spectrum plan report) (2010)
26. Office of Communications: Spectrum usage rights technology and usage neutral access to the radio spectrum (2006)
27. Opinion on wireless access policy for electronic communications services (WAPECS) (a more flexible spectrum management approach). Radio Spectrum Policy Group. RSPG05-102
28. Peha, JM.: Sharing spectrum through spectrum policy reform and cognitive radio. Proc. IEEE **97**(4), 708–719 (2009)
29. President's Council of Advisors on Science and Technology: Realizing the full potential of government-held spectrum to spur economic growth, July 2012
30. Radiocommunication Sector of the International Telecommunications Union: Final acts, world radiocommunication conference 2012 (WRC-12), 2012
31. Radio Spectrum Policy Group: Opinion on cognitive technologies. RSPG10-348 (2011)
32. Report on collective use of spectrum (CUS) and other spectrum sharing approaches. Radio Spectrum Policy Group. RSPG11-392

Chapter 11
Simulation of Cognitive Radio Networks in OMNeT++

Giuseppe Caso, Luca De Nardis, and Oliver Holland

Abstract A widespread methodology for performance analysis and evaluation in communication systems engineering is network simulation. It is widely used for the development of new architectures and protocols. Network simulators allow to model a system by specifying both the behavior of the network nodes and the communication channels, and Cognitive Radio (CR)-related research activities have been often validated and evaluated through simulation too.

Following this approach, this chapter presents an ongoing effort towards the development of a CR simulation framework, to be used in the design and the evaluation of protocols and algorithms. OMNeT++, in combination with MiXiM framework functionalities, was chosen as the developing platform, thanks to its open source nature, the existing documentation on its architecture and features, and the user-friendly Integrated Development Environment (IDE).

11.1 Introduction

The term Cognitive Radio (CR) is often used as a synonym for Dynamic Spectrum Access (DSA): it refers to a novel approach to wireless communications, aiming at improving the performance of a radio network by introducing *cognition*. A CR is a context-aware radio capable of autonomous reconfiguration (in the form

G. Caso (✉) • L. De Nardis
DIET Department, Sapienza University of Rome, Rome, Italy
e-mail: caso@newyork.ing.uniroma1.it; lucadn@newyork.ing.uniroma1.it

O. Holland
Institute of Telecommunications, King's College London, London, UK
e-mail: oliver.holland@kcl.ac.uk

of a Software Defined Radio – SDR) by learning from and adapting to the communication environment. The main goal in cognitive radio research is to design communications networks that (cooperatively or not) coexist with other networks, by avoiding mutual interference and at the same time efficiently using the available frequency spectrum.

Until recently, research about CR mainly focused on a peculiar cognitive function called Spectrum Sensing [1]. Through Spectrum Sensing (SS), a CR tries to detect the presence/absence of other users (from different networks) on a investigated frequency band. As a result of SS, each user in a Cognitive Radio Network (CRN) takes a local decision on the channel state, variable depending from both space and time. In order to obtain a global network decision, several works proposed Cooperative Spectrum Sensing (CSS) schemes, where the nodes in the CRN, or a subset of them, share the results of sensing among them or to a central unit [2]. CSS can be very important in particular scenarios when detection of the Primary Users (PUs) of a particular band is transmitter-based and the PU receiver is located in the Secondary Users (SUs – the Cognitive Radios) transmission range (Hidden PU problem). An increased probability of correct decision on the spectrum usage can be obtained by cooperation and information exchange between CRs.

Network simulators are widely used for the development of new architectures and protocols, allowing to model a system by specifying both the behavior of the network nodes and the communication channels. The implementation of a CR simulator could indeed prove very useful in the study of CR networks, in particular for the evaluation of cognitive protocols and algorithms. Moving from this observation, efforts were dedicated to the creation of such a simulator, with the final goal of creating a software tool capable of supporting the definition and performance evaluation of cooperative spectrum sensing algorithms and, more in general, of solutions for the coexistence of heterogeneous systems operating at the same frequency bands, by setting realistic application scenarios with proper parameters and protocols of the SUs forming the CRN. The platform selected as the basis for the development of the CR simulator is formed by the OMNeT++ open source network simulator [3], extended with the functionalities provided by the MiXiM framework [4]. OMNeT++/MiXiM was a natural choice, due to its open source nature, the availability of large and accurate documentation, and the user-friendly Integrated Development Environment (IDE) based on Eclipse.

This chapter presents early results obtained in the implementation of the CR simulator and its use for addressing open scientific issues related to the deployment of CR networks.

In the next section, the theoretical model driving the implementation of the simulator is given. Then, in the following sections, brief reviews of OMNeT++ software and MiXiM project are given, in order to provide a clear view of the framework within which the CR simulator is being developed. The last section presents a set of examples of the problems currently being addressed with the aid of the simulator.

11.2 System Model

Spectrum Sensing is a fundamental function for a CR device. Several SS techniques were proposed in the last years. Sensing based on energy detection (ED-SS) is widely adopted thanks to its low costs and complexity. In ED-SS, CRs do not have knowledge on the PUs signal; they measure the energy of the received waveform in a given channel, over an observation time window of T (seconds). Next, they compare the test statistic Y, approximating the signal energy in the interval $(0, T)$, with a threshold λ, whose optimum value depends on the noise floor [5]: if the evaluated energy is higher (respectively lower) than the threshold, then SU decides for PU presence (respectively absence). Defining the two hypotheses with $\mathcal{H}_0$ and $\mathcal{H}_1$, the decision problem can be expressed as follows:

$$\mathcal{H}_0 : Y < \lambda,$$
$$\mathcal{H}_1 : Y \geq \lambda.$$

In a Local Spectrum Sensing (LSS) scenario, a CR node opportunistically transmits when it does not detect presence of any PUs, and its decision is not related to SS results of other SUs in the surrounding environment.

In a non-fading environment, denoting with γ the PU signal-to-noise ratio (SNR) at the SUs within a channel of bandwidth W (Hertz) and assuming for the test statistic Y, in hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, central and non-central (with parameter 2γ) chi-square distributions with $2TW$ degrees of freedom, probability of correct detection P_d, and probability of false alarm P_fa, are as follows:

$$P_\mathrm{d} = P\{Y > \lambda | \mathcal{H}_1\} = Q_m(\sqrt{2\gamma}, \sqrt{\lambda}), \tag{11.1}$$

$$P_\mathrm{fa} = P\{Y > \lambda | \mathcal{H}_0\} = \frac{\Gamma(m, \frac{\lambda}{2})}{\Gamma(m)}, \tag{11.2}$$

where it is assumed that Time-Bandwidth product TW is the integer number m, $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ are the complete and incomplete gamma functions, and $Q_m(\cdot, \cdot)$ is the generalized Marcum Q-function, defined from the $I_{m-1}(\cdot)$ modified Bessel function of $(m-1)$th order [2].

For large values of m, the *Gaussian Approximation* can be applied to the test statistic Y under either $\mathcal{H}_0$ or $\mathcal{H}_1$ [5]. Under the $\mathcal{H}_0$, Y is the sum of $2m$ statistically independent random variables. Therefore, since $\mathbb{E}[Y] = 2m$ and $\mathsf{Var}[Y] = 4m$, Y is distributed as a Gaussian random variable denoted by $\mathcal{N}(2m, 4m)$, and the P_fa is given by:

$$P_\mathrm{fa} = \frac{1}{\sqrt{8\pi m}} \int_Y^\infty e^{-\frac{(x-2m)^2}{8m}} \, dx = \frac{1}{2}\mathrm{erfc}\left(\frac{\lambda - 2m}{2\sqrt{2m}}\right). \tag{11.3}$$

Under $\mathcal{H}_1$, $\mathbb{E}[Y] = 2m + 2\gamma$ and $\mathsf{Var}[Y] = 4(m + 2\gamma)$, and therefore $Y \sim \mathcal{N}(2m + 2\gamma, 4(m + 2\gamma))$. P_d is given by:

$$P_\mathsf{d} = \frac{1}{2}\mathrm{erfc}\left(\frac{\lambda - 2m - 2\gamma}{2\sqrt{2}\sqrt{m + 2\gamma}}\right). \tag{11.4}$$

When the channel gain h is varying due to shadowing/fading, (11.1) is conditioned on the instantaneous γ. In this case, P_d is derived by averaging (11.1) over fading statistics:

$$P_\mathsf{d} = \int_\gamma Q_m(\sqrt{2\gamma}, \sqrt{\lambda}) f_\Gamma(\gamma)\, d\gamma, \tag{11.5}$$

where $f_\Gamma(\gamma)$ is the pdf of SNR under fading.

Cooperative Spectrum Sensing aims at improving LSS performance. In hard decision distributed CSS, each CR node takes and shares with its neighbors an independent decision and then it applies a fusion rule on them. The generic hard fusion rule is the *k out of n* rule: if k or more nodes decide the hypotheses $\mathcal{H}_1$, then the node will decide for $\mathcal{H}_1$. When $k = 1$, the rule becomes the OR rule; when $k = n$ the fusion rule works as the AND rule; when $k = (n + 1)/2$, the fusion rule becomes the Majority rule.

Let N be the number of cooperative SUs, experiencing independent and identically distributed fading/shadowing with same average SNR. The SUs employ ED and use the same decision rule (with threshold λ). If a SU receives decisions from $N - 1$ users and it applies the generic n-out-of-N, then the probabilities of detection and false-alarm for the collaborative scheme (Q_d and Q_fa, respectively) are [2]:

$$Q_\mathsf{d} = \sum_{k=n}^{N} \binom{N}{k} P_\mathsf{d}^k (1 - P_\mathsf{d})^{N-k}, \tag{11.6}$$

$$Q_\mathsf{fa} = \sum_{k=n}^{N} \binom{N}{k} P_\mathsf{fa}^k (1 - P_\mathsf{fa})^{N-k}, \tag{11.7}$$

where P_d and P_fa are the individual probabilities of detection and false alarm as defined before. Using the OR rule (1 out of N), (11.6) and (11.7) become:

$$Q_\mathsf{d} = 1 - (1 - P_\mathsf{d})^N, \tag{11.8}$$

$$Q_\mathsf{fa} = 1 - (1 - P_\mathsf{fa})^N. \tag{11.9}$$

Formulas for Majority rule ($\lceil N/2 \rceil$ out of N) are:

$$Q_\mathsf{d} = \sum_{k=\lceil N/2 \rceil}^{N} \binom{N}{k} P_\mathsf{d}^k (1 - P_\mathsf{d})^{N-k}, \tag{11.10}$$

$$Q_{\text{fa}} = \sum_{k=\lceil N/2 \rceil}^{N} \binom{N}{k} P_{\text{fa}}^k (1 - P_{\text{fa}})^{N-k}. \tag{11.11}$$

Lastly, for the AND (N out of N), one obtains:

$$Q_{\text{d}} = P_{\text{d}}^N, \tag{11.12}$$

$$Q_{\text{fa}} = P_{\text{fa}}^N. \tag{11.13}$$

When the network is infrastructure based, there will be a Base Station (BS), also known as Fusion Center (FC), for CSS purposes. The BS will be the device in the network that will apply the fusion rule and then will broadcast the cooperative decision. If the channels between SUs and FC are perfect, Q_{fa} and Q_{d} are defined as before.

11.2.1 Operating Mode 1: Constant False Alarm Rate (CFAR)

In this scenario, the overall CRN has fixed a target probability of false alarm $\bar{Q}_{\text{fa}}$, in order to optimize the usage of spectrum opportunities. Given $\bar{Q}_{\text{fa}}$, the corresponding $\bar{P}_{\text{fa}}$ can be obtained inverting the chosen fusion rule formula. This leads to the evaluation of the threshold λ, inverting (11.3), and the consequent evaluations of P_{d} and Q_{d}, for a given value of γ. The generic formulation of the threshold λ is as follows:

$$\lambda^{\text{CFAR}} = \text{erfc}^{-1}(2\bar{P}_{\text{fa}})[2\sqrt{2m}] + 2m. \tag{11.14}$$

In the case of OR and AND rules, it is easy to invert the fusion rule formula. In the case of Majority rule, an approximation to solve the issue was proposed in [6] allowing to obtain a good estimate of $\bar{P}_{\text{fa}}$ without inverting (11.11).

11.2.1.1 Hard Decision Fusion Rules

OR/AND rules – Inverting respectively (11.9) and (11.13), one obtains:

$$\bar{P}_{\text{fa}} = 1 - \sqrt[N]{1 - \bar{Q}_{\text{fa}}} \quad \text{OR rule,} \tag{11.15}$$

$$\bar{P}_{\text{fa}} = \sqrt[N]{\bar{Q}_{\text{fa}}} \qquad \text{AND rule.} \tag{11.16}$$

At this point, it is possible to compute λ^{CFAR} using Eq. (11.14) in order to subsequently obtain P_{d} and Q_{d}.

Majority rule – Instead of inverting (11.11), an approximation of it is proposed. By applying the de Moivre-Laplace theorem to the argument of (11.11), one obtains:

$$\binom{N}{k} \bar{P}_{\text{fa}}^k (1 - \bar{P}_{\text{fa}})^{N-k} \approx \frac{1}{\sqrt{2\pi N \bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}} e^{\frac{-(k - N\bar{P}_{\text{fa}})^2}{2N\bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}}. \tag{11.17}$$

Using (11.17) in (11.11), one has

$$\bar{Q}_{\text{fa}} \approx \sum_{k=\lceil N/2 \rceil}^{N} \frac{1}{\sqrt{2\pi N \bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}} e^{\frac{-(k - N\bar{P}_{\text{fa}})^2}{2N\bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}}. \tag{11.18}$$

For $N \to \infty$, the summation, same for a multiplicative factor, converges to the Riemann integral of the argument. One can indeed introduce the integral but, in the approximation, the upper limit is ∞, introducing thus an additional term to the summation: this term is an erfc(.), converging to zero when $N \to \infty$. For this reason, given a finite value of N, the summation is well matched with the integral and the additional term is negligible:

$$\left| \sum_{k=\lceil N/2 \rceil}^{N} \frac{1}{\sqrt{2\pi N \bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}} e^{\frac{-(k - N\bar{P}_{\text{fa}})^2}{2N\bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}} - \int_a^{\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx \right| \leqslant \delta(N), \quad (11.19)$$

where $a = \frac{\lceil N/2 \rceil - \frac{1}{2} - N\bar{P}_{\text{fa}}}{\sqrt{2N\bar{P}_{\text{fa}}(1 - \bar{P}_{\text{fa}})}}$ and $\delta(N)$ is a positive number decreasing with N. Finally, one can write:

$$\left| \sum_{k=\lceil N/2 \rceil}^{N} \binom{N}{k} \bar{P}_{\text{fa}}^k (1 - \bar{P}_{\text{fa}})^{N-k} - \frac{1}{2}\text{erfc}(a) \right| \leqslant \varepsilon(N), \tag{11.20}$$

where $\varepsilon(N)$ is, again, a positive number decreasing with N, representing the discrepancy between theory and approximation; Q_{fa} can be thus approximated by using erfc(a) function.

Figure 11.1 compares the $\bar{Q}_{\text{fa}}$ obtained with OR, AND, exact Majority formula, and the proposed approximation, showing the good match between exact and approximated Majority formulas; the approximation can be used to easily get the $\bar{P}_{\text{fa}}$ required to obtain a target $\bar{Q}_{\text{fa}}$. Figure 11.2 presents the value of ε as a function of the number of SUs; the approximated formula gets more and more accurate as the number of SUs increases.

11.2.2   Operating Mode 2: Constant Detection Rate (CDR)

In CDR mode it is assumed that a target $\bar{Q}_{\text{d}}$ was selected for the CRN. Given $\bar{Q}_{\text{d}}$, the corresponding $\bar{P}_{\text{d}}$, is obtained by inverting the formula of the selected fusion
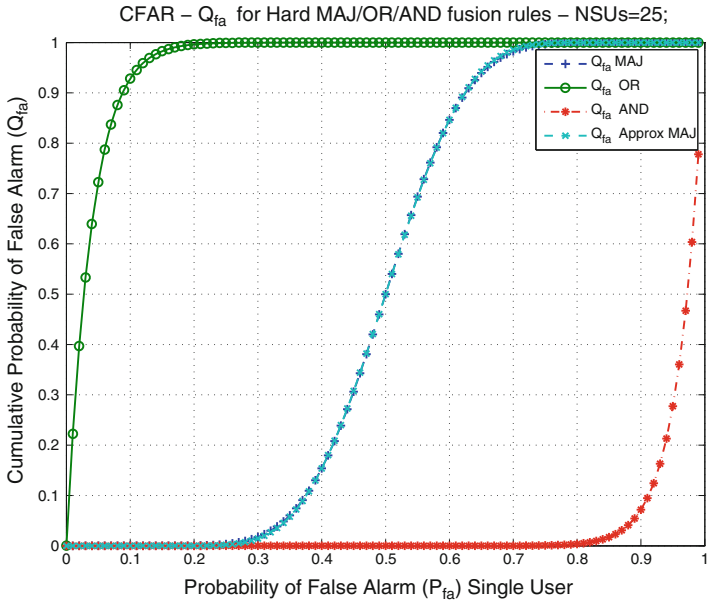
Fig. 11.1 CFAR – target $\bar{Q}_{fa}$ vs. target $\bar{P}_{fa}$
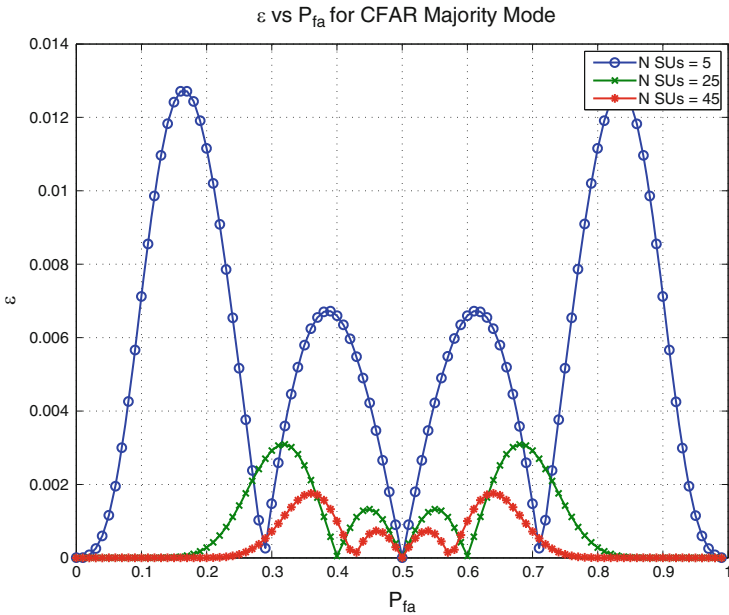
Fig. 11.2 Theoretical vs. approximated formula for $\bar{Q}_{fa}$ in terms of $\varepsilon(N)$ for CFAR majority operating mode
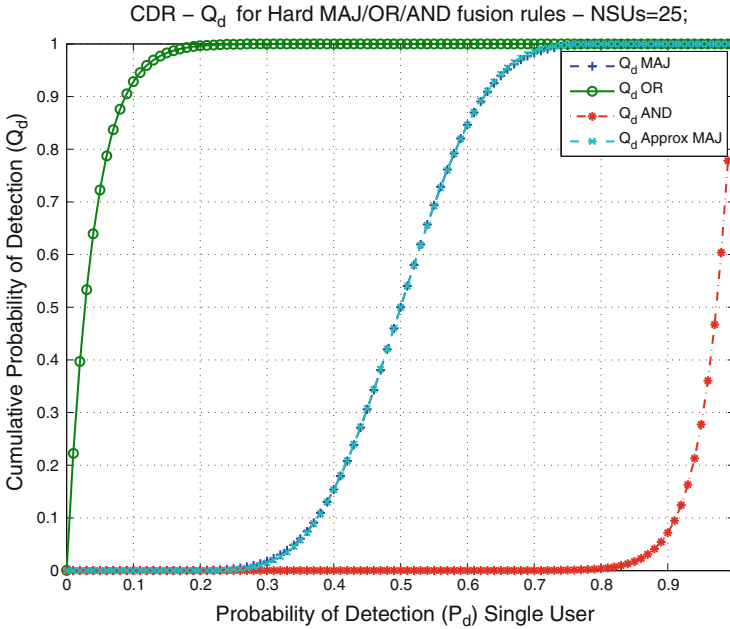
Fig. 11.3 CDR – target $\bar{Q}_d$ vs. target $\bar{P}_d$

rule. This leads to the evaluation of the threshold λ, by inverting Eq. (11.4) for a given value of γ, and the consequent evaluations of P_{fa} and Q_{fa}. λ is as follows:

$$\lambda^{CDR} = \text{erfc}^{-1}(2\bar{P}_d)[2\sqrt{2(m + 2\gamma)}] + 2(m + \gamma). \qquad (11.21)$$

11.2.2.1 Hard Decision Fusion Rules

OR/AND rules – Inverting respectively (11.8) and (11.12), one obtains:

$$\bar{P}_d = 1 - \sqrt[N]{1 - \bar{Q}_d} \quad \text{OR rule}, \qquad (11.22)$$

$$\bar{P}_d = \sqrt[N]{\bar{Q}_d} \qquad \text{AND rule}. \qquad (11.23)$$

It is possible now to compute λ^{CDR} using Eq. (11.21) in order to subsequently obtain P_{fa} and Q_{fa}.

Majority rule – In this case an approximation for (11.10) is proposed. Following the same steps as in Sect. 11.2.1, the following inequality is obtained [6]:

$$\left| \sum_{k=\lceil N/2 \rceil}^{N} \binom{N}{k} \bar{P}_{\mathsf{d}}^{k} (1 - \bar{P}_{\mathsf{d}})^{N-k} - \frac{1}{2}\mathrm{erfc}(a) \right| \leqslant \varepsilon(N). \qquad (11.24)$$

Figure 11.3 shows the match between approximated and exact formulas for $\bar{Q}_{\mathsf{d}}$ as a function of $\bar{P}_{\mathsf{d}}$. The role of the number of SUs on the accuracy of the approximation is comparable to the CFAR case discussed in Fig. 11.2.

11.3 OMNeT++: Objective Modular Network Testbed in C++

OMNeT++ (Objective Modular Network Testbed in C++) is a popular tool for discrete event simulations [3]. This simulator is widely used, as example for: traffic modeling of telecommunication networks, protocols modeling, modeling queuing networks, modeling multiprocessors and other distributed hardware and, in general, modeling of scenarios when several entities behave independently and the output of every entity must be analyzed. OMNeT++ allows detailed simulations of the effects of independent communications and provides a framework and the matching infrastructure for writing the C++ code required for the simulations. Specific application areas are covered by various open source simulation frameworks. These models are developed independently of OMNeT++: as example, the Mobility Framework, used for mobile and wireless simulations, the INET Framework, for wired and wireless TCP/IP based simulations, Castalia, particularly for wireless sensor networks and MiXiM, used for mobile and wireless simulations.

OMNeT++ provides a C++ class library and utility classes, in order to create simulation components (modules and channels); proper language to design simulations and configure them (NED language, .ini files); user interfaces and environments for simulations (Tkenv, Cmdenv); an Eclipse-based IDE for designing, executing and evaluating the projects; interfaces for real-time and, eventually, parallel simulations, and so on. One of the main feature is the reusable form of the OMNeT++ modules, so that the OMNeT++ models are assembled from them, combined and connected with each other via gates, in order to form so-called compound modules. A simple module, for example, can be connected with another simple module, or with a compound module. Basically, NED files in the projects define the model structure (topology), and leave undefined the behaviour and some modules parameters. Accordingly, behaviour is defined via C++ code, subclassing the basic simple modules, and parameters not assigned in NED files are set in the .ini files.

Building and compiling the simulation program is usually quite simple. In the most simple case, where all files are in one project directory, one only needs to create a makefile with the appropriate settings. If the simulation sources are spread over several directories (typically when one or more frameworks like MiXiM are used), the simulation needs to link with code from other directories, by passing additional options. To run the executable, at least an .ini file is needed, in order to set the network, to specify the NED files paths, to assign open parameters, the

simulation time duration, the seeds for random numbers, the results collection, and any other settings for the specific experiment.

To summarize, in OMNeT++, C++ and the NED files represent the model, and experiments are described in .ini files that let to keep the model unchanged while exploring the variability of the parameters, e.g. by varying the number of nodes in the network. Simulation results are recorded into output vector (.vec) and output scalar (.sca) files and this functionality has to be programmed into the simple modules. An output vector file contains series of (timestamp, value) pairs in order to store the variability of some parameters over time (as examples, queue lengths, end-to-end delays of received packets, channels throughput); output scalar files contain summary statistics: number of packets sent/dropped, average end-to-end delay of received packets, peak throughput. Simulation results can be visualized in the Analysis Editor of the OMNeT++ IDE.

11.4   MiXiM: The Mixed Simulator

MiXiM (MIXed sIMulator) is the result of merging of several OMNeT++ frameworks written to support mobile and wireless simulations: ChSim by Universitaet Paderborn, Mac Simulator by Technische Universiteit Delft, Mobility Framework by Technische Universitaet Berlin and Positif Framework by Technische Universiteit Delft, among others [4]. It provides, in a user-friendly graphical representation, detailed models of the wireless propagation channel, mobility and obstacles models, and many communication protocols, especially at the MAC level. Models and protocols provided by MiXiM can be divided into five groups:

- *Environment Models* – To take into account relevant parts of the real world, such as obstacles for wireless communications;
- *Connectivity and Mobility* – To take into account mobility of objects and nodes influencing packets transmission/reception;
- *Reception and Collision* – To model how a transmitted signal changes on its way to the receivers, because of the interference from other senders;
- *Experiment Support* – To compare the results with an ideal state and to support different evaluation methods;
- *Protocol Library* – A rich protocol library enabling detailed simulation scenarios and comparisons with already implemented ones.

MiXiM supports the simulation of networks with a very high number of nodes with low memory and execution time costs.

The structure of the framework can be divided into two parts: the Base Framework and the Protocol Library. The Base Framework provides the general wireless simulations functionalities, such as mobility and propagation channel and communications models; the Protocol Library integrates the Base Framework with a large set of standard protocols. Modules reusability makes easy to implement new protocols, allowing every module to be replaced by another module, adding new functionalities
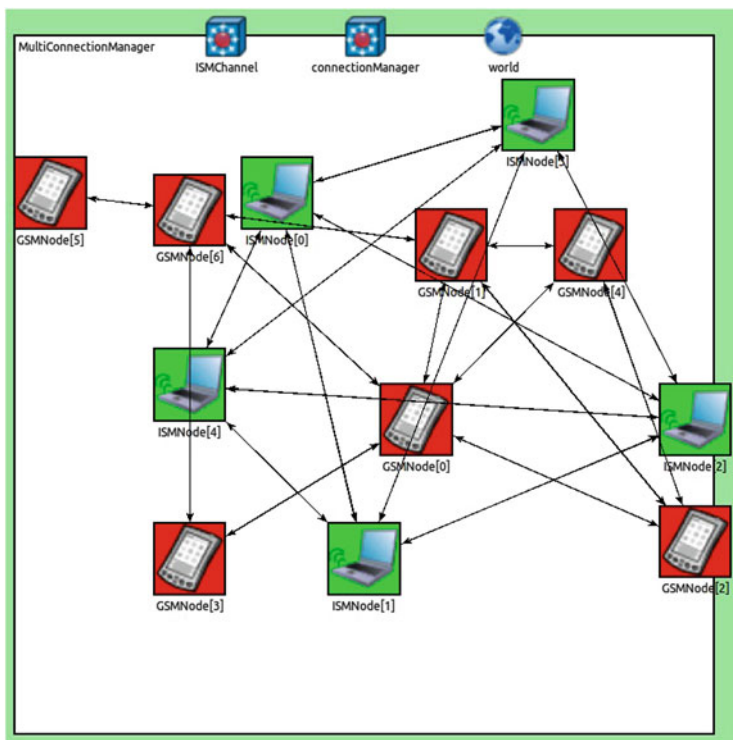
Fig. 11.4 Example of MiXiM simulation environment

to the base implementation. For some of these modules there is already a wide choice of protocols available: in terms of MAC protocols, MiXiM already contains standard MAC protocols, as, for example, the IEEE 802.11b/g family, as well as the IEEE 802.15.4 standard. For the Network Layer protocols, MiXiM supports networking protocols for several traffic paradigms (source-to-sink, any-to-any, local neighborhood, etc.). Moreover, regarding the Mobility models, MiXiM already presents a rich library of mobility modules, which includes simple modules like "Constant Speed Mobility" and "Circle Mobility". It is also very easy to create new mobility modules, by sub-classing the "BaseMobility" class that provides all the basic functionalities for a mobile node. For this reason, only the specific mobility pattern has to be implemented in order to create a new mobility module.

Generically, a MiXiM-based simulator describes a network with a parameterizable number of nodes eventually moving in a playground with or without obstacles. Each host has a defined transmission power and is basically an OMNeT++ compound module of simple modules matching the layers defined in the generic ISO/OSI stack.

Figures 11.4 and 11.5 present an example of MiXiM simulation network, of host structure and of NIC module (compound module containing MAC layer and Physical layer simple modules), respectively.
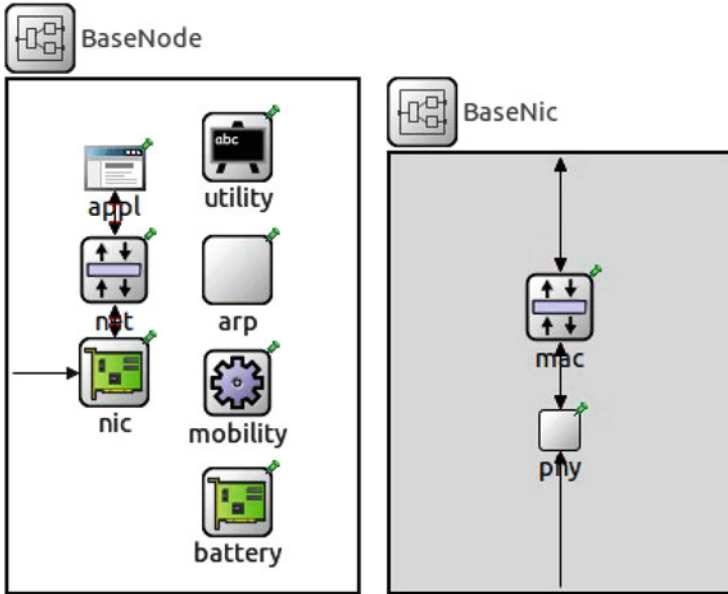
Fig. 11.5 Example of MiXiM base host: BaseNode architecture (on the *left*), NIC architecture (on the *right*)

An exhaustive explanation of MiXiM methods and modes of operation is out of the scope of this work. The interested reader can refer to [4] as well as to the official MiXiM website [7] for a generic overview and presentation of the framework. For a complete architectural review, particularly focused on the Physical layer, one can refer to [8].

11.4.1 From a MiXiM-Based to a Cognitive Radio Scenario

The design of the CRN scenario, as presented in the System Model paragraph, started from a generic MiXiM scenario extended with fundamental functionalities simulating peculiarities of the Primary User vs. Secondary and Cognitive Users case study:

- Introduction of a specific Module simulating the Primary User;
- Introduction of the Secondary Network, addressing the main differences between Base Station Node and Common Nodes (SUs);
- Introduction of peculiar cognitive functionalities in the OSI stack of a generic MiXiM host (particularly, the Sensing and, if required, the Clustering capabilities);
- Introduction of Multi-Channel case;

  A typical simulation scenario is presented in Fig. 11.6.
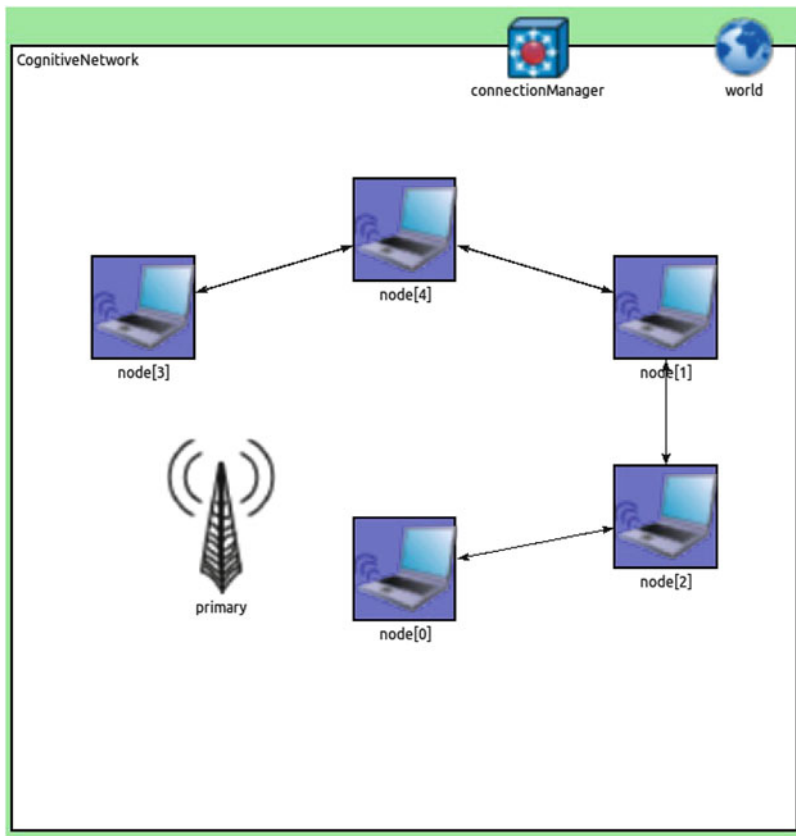
Fig. 11.6 Example of MiXiM-based cognitive scenario

Focusing in particular on the structure of the SU, this was obtained by extending the base MiXiM host, introducing the following functionalities:

- *Sensing and Clustering Modules*: These two modules were added to fulfill, respectively, to the spectrum sensing and the clustering functions;
- *Multi-Channel IS Multi-NIC*: The architectural choice to create the possibility for the hosts to transmit (and, for the BS, to receive) on different channels was to equip the hosts with multiple single-channel NIC (specifically, a NICs array), each one defined on a different carrier frequency and working on a different channel. A common dedicate channel for sensing and control info exchange between SUs and the BS is assumed;
- *NIC module*: Each NIC module inside each host contains two simple modules, used to define the MAC and the Physical layers. The used NICs are 802.11b NIC, already implemented in the MiXiM library, working on different 802.11b channels (or, with appropriate changes, on DVB-T channels). For this reason the
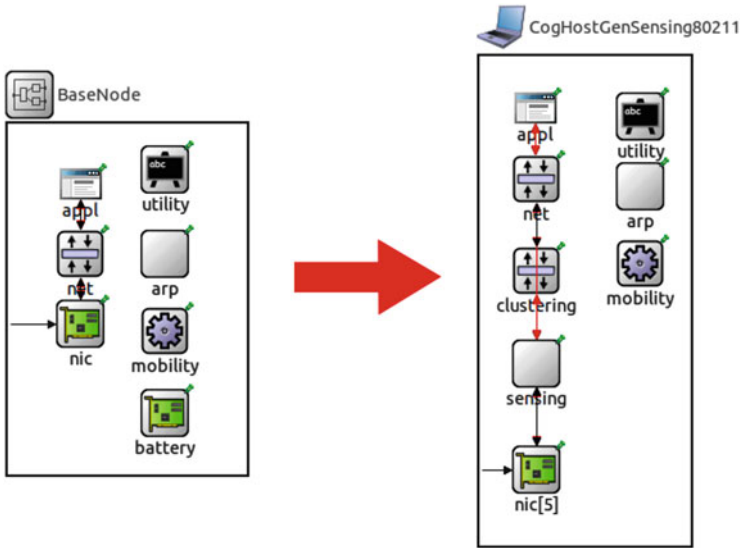
Fig. 11.7 Architecture of a cognitive host compared to the architecture of the *BaseNode* provided by MiXim

MAC layer is substantially based on the CSMA/CA multiple access technique and the assumption is to have 802.11-like CR devices.

Figure 11.7 shows the internal SU architecture, highlighting, how the *BaseNode* module defined in MiXiM was extended to include functionalities typical of a Cognitive Radio device.

11.5 Application of the CR Simulator to Open Research Issues

This section provides two examples of how it is possible to use the CR simulator in order to study the performance of MAC and network protocols in specific application scenarios by adjusting parameters as PUs behavior, cognitive network features (number of SUs, Energy Detector-based Spectrum Sensing modes, mobility models), operating frequencies (e.g. frequencies used by DVB-T signals vs. Wi-Fi), medium access modalities, propagation channel models, network types and topologies, and so on. The examples presented in the rest of this section deal with the two following topics:

- *Cooperative Spectrum Sensing based on Hard Decision fusion rules under CFAR and CDR constraints*;

- *Impact of Spatio-Temporal Correlation in Cooperative Spectrum Sensing for Mobile Cognitive Radio Networks.*

As detailed in the following, the CR simulator allowed to carry out an extensive comparative analysis regarding various aspects defining a CRN, in particular, (1) spectrum sensing performance, both local and centralized cooperative with different hard decision fusion rules, (2) impact of mobility, (3) impact of channel spatio-temporal correlation, (4) introduction of cluster-based solutions for the network organization and nodes selection, (5) achievable data throughput for the cognitive network.

11.5.1 Cooperative Spectrum Sensing Based on Hard Decision Fusion Rules Under CFAR and CDR Constraints

This section summarizes the study reported in [6], focusing on the performance of Cooperative Spectrum Sensing, assuming the theoretical system model introduced in Sect. 11.2.

11.5.1.1 Simulation Settings

The scenario considered in [6] foresees a PU located in the top left corner of a square area of 20×20 km^2, transmitting with a fixed transmitter power (200 kW) in a single DVB-T 8 MHz channel in the UHF bandwidth. A CRN operates within the coverage area of the PU, formed by SUs located at the lower right part of the playground, within a 700×700 m^2 area. A Fusion Center is positioned in the center of such area, and SUs communicate among them and with the FC using with a transmission power of 110 mW. A control channel used to exchange generic control packets and sensing information is also simulated.

Each simulation run covers 1 h of simulated time, during which each SU takes a local decision within a sensing phase of 50 μs and then transmits it to the FC during the subsequent exchange phase of 1 s. The FC takes a global decision every 5 s.

11.5.1.2 Results

Figures 11.8 and 11.9 show achievable Q_d (CFAR mode) and Q_{fa} (CDR mode), respectively as a function of the requested targets and for different values of γ and using a Majority rule in the FC. Perfect knowledge about the number of collaborating SUs ($N = 25$) and the average SNR ($\gamma = [0:5:30]$ dB), is assumed, leading in turn to optimal choice of ED threshold λ. As a result, all targets are met for both the operating modes. Moreover, it is demonstrated that for low γ values, the CRN should not set too stringent targets as this actually leads to worse
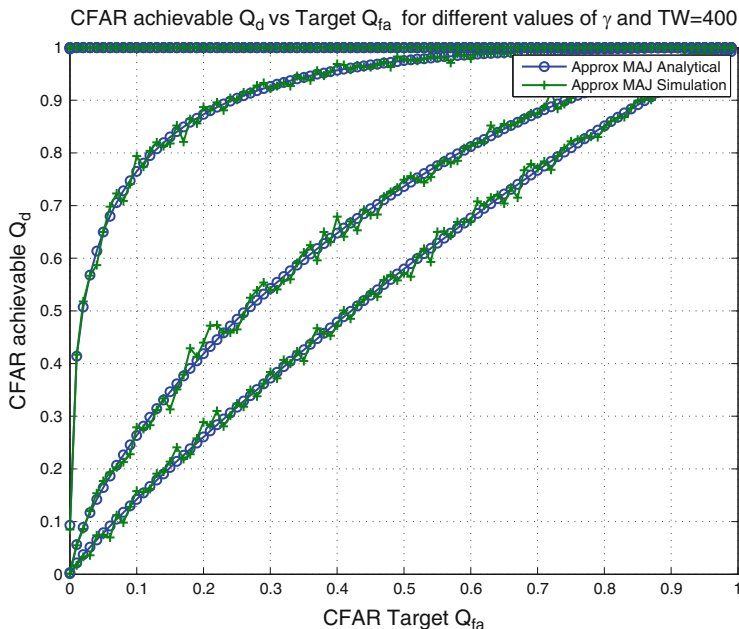
Fig. 11.8 CFAR – achievable Q_d vs. target $\bar{Q}_{fa}$
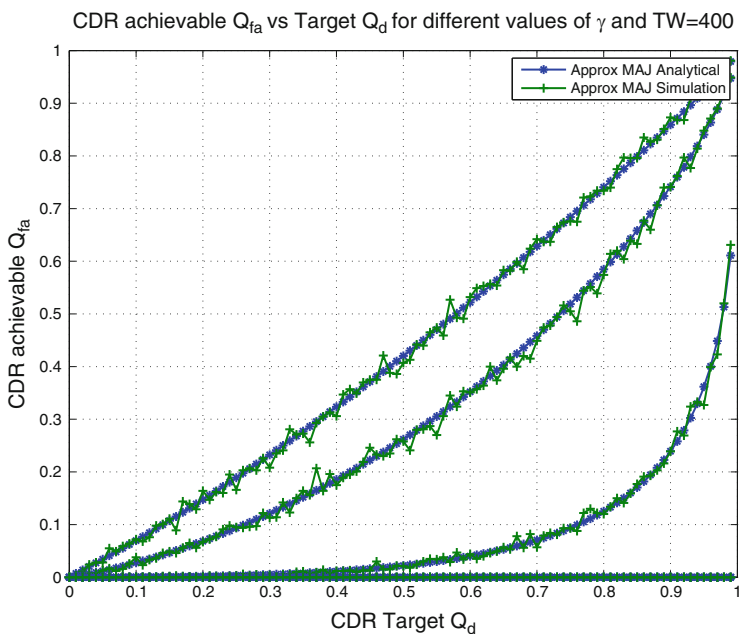
Fig. 11.9 CDR – achievable Q_{fa} vs. target $\bar{Q}_d$
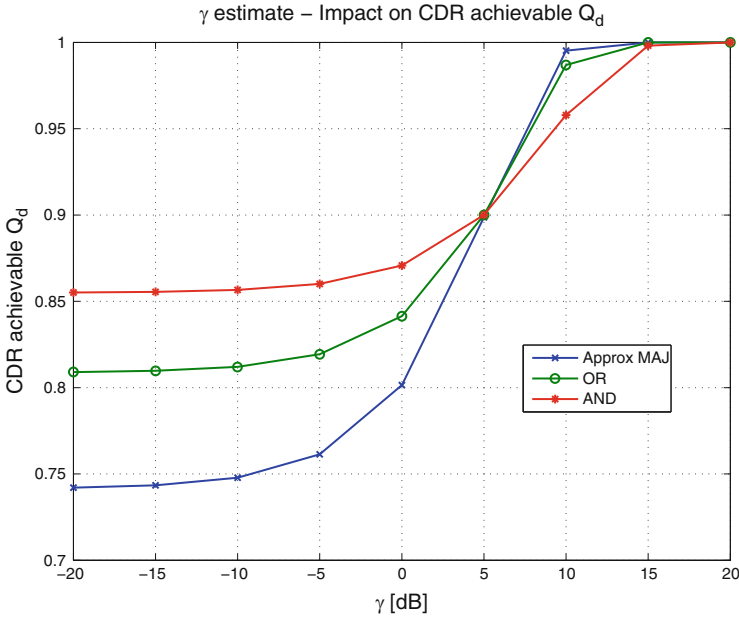
Fig. 11.10 CDR achievable Q_d – ED-SS threshold λ is fixed in order to achieve $\bar{Q}_d = 0.9$ if $\bar{\gamma} = 5\,\mathrm{dB}$

global cooperative performance: figures show that a low $\bar{Q}_{fa}$ leads to poor detection performance for CFAR, while high $\bar{Q}_d$ lead to unacceptable false alarm rates for CDR.

The number of cooperating SUs and the expected average PU-related SNR γ sensed by the CRN are required to define the threshold λ. As regards N, a reasonably good estimate can be obtained by relying on a *training* phase involving SUs and FC. Estimating γ is however much harder with the assumption of no a-priori knowledge about the PU. Focusing on CDR mode, Fig. 11.10 shows the impact of errors in γ estimate on the target $\bar{Q}_d$ achievability, assuming $\bar{Q}_d = 0.9$, $N = 25$ and $\bar{\gamma} = 5\,\mathrm{dB}$. As expected, for values of γ lower than $\bar{\gamma}$, the CRN is not able to meet the target $\bar{Q}_d$. The floor observed for the lowest value of γ is due to the hypothesis of continuous activity of the PU (that allows the evaluation of Q_d even if in a *very low-γ* case). Moreover, the inequality $\lambda_{MAJ} > \lambda_{OR} > \lambda_{AND}$ is verified and this explains why the Majority rule behaves as a lower (upper) bound when the λ estimate is too optimistic (pessimistic).

In [6], CSS schemes under CFAR and CDR constraints were analyzed, adopting OR, AND and Majority decision rules. Novel approximations for the network probability of false alarm (CFAR) and detection (CDR) as a function of individual probabilities were proposed and validated by means of simulations, and the role of errors in the estimation of parameters used for setting individual performance requirements was investigated.

11.5.2 Impact of Spatio-Temporal Correlation in Cooperative Spectrum Sensing for Mobile Cognitive Radio Networks

This section presents a second example of application of the CR simulator to the study of CRNs, summarizing the study reported in [9], where a more realistic evaluation of the impact on CSS performance of mobility and spatio-temporal correlation between SUs measurements was proposed, by relaxing and removing some of the assumptions found in previous works on this topic, by simulating realistic conditions for channel correlation and SUs mobility. The work addresses the problem by defining a node selection metric based on the statistical index known as Moran's I [10]. In the proposed framework the Moran's I is used to determine the degree of correlation between decisions taken by different SUs in different locations of the defined environment, and to select a sub-optimal group of quasi-uncorrelated SUs to be involved in the CSS procedure.

11.5.2.1 Simulation Settings

The scenario considered in [9] is similar to the one already described in Sect. 11.5.1.1, with the following differences:

- The total area size is $20 \times 20 \, \text{km}^2$;
- SU mobility is taken into account in the analysis by having the SUs moving within the working area using a Gauss-Markov mobility model with the following average speed values: $v = \{5, 10, 15, 20\} \, \text{m/s}$;
- The CRN playground is divided in $16 \, 175 \times 175 \, \text{m}^2$ square cells, used to evaluate the Moran's I index.

11.5.2.2 Results

Figure 11.11 presents the impact on nodes selection, showing the average number of SUs collaborating in the CSS during the simulation, for both static and mobile cases.

Until the number of SUs is *low* (respect to the number of cells in which the playground was divided for Moran's I evaluation), practically no nodes selection occurs, due to the hypotheses of randomly chosen SUs positions in the static scenario and *quasi*-random feature of the mobility model that, on average, lead to a low degree of correlation (on average, the SUs are spatially dispersed in the playground). Finally, when the number of SUs is higher that the number of cells, a higher number of SUs in the network, corresponds to a higher the number of discarded SUs. This is actually related to the obvious direct proportionality between the number of SUs and the degree of correlation of the SUs' decisions.
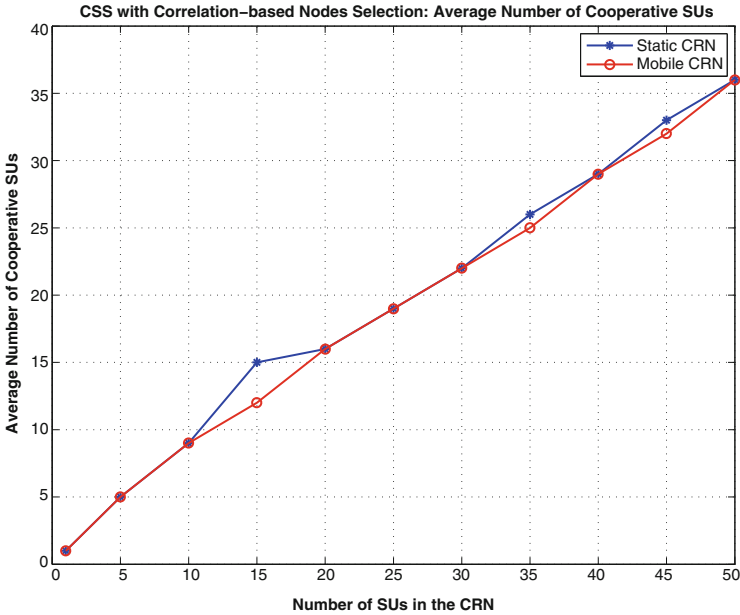
Fig. 11.11 Average number of cooperative SUs for CSS with proposed 'Node Selection' scheme

Figures 11.12 and 11.13 show, respectively, the measured Q_d for CSS with Majority rule, as a function of the CFAR target $\bar{Q}_{fa}$ and the number of SUs ($N = \begin{bmatrix} 1 & 5 & 15 & 25 & 35 & 45 \end{bmatrix}$), for schemes without and with nodes selection. Similar results in Figs. 11.14 and 11.15 for, respectively, AND and OR rules with $N = \begin{bmatrix} 1 & 5 & 50 \end{bmatrix}$. For the evaluation of the single user P_d an average $\gamma = 5\,\text{dB}$ is assumed.

The plots for Majority rule show that, after a significant improvement given by cooperation of SUs, the performance does not improve significantly with the number of SUs, making the use of more SUs less and less useful. Therefore, from this point of view, the scheme with nodes selection achieves comparable performance with respect to the previous scheme even if with a lower number of cooperative SUs. Similar results were obtained for AND and OR rules.

Finally, Fig. 11.16 presents the performance of CSS with Majority fusion rule, for the scheme without nodes selection and a number of SUs belonging to the network equal to 25. Differently from the previous analysis, in this case a Gauss-Markov mobility model is assumed for each SU. The results do not show the clear improvement shown in [11, 12] for a single mobile SU and this is probably related to the fact that a realistic modeling of propagation channel and mobility behaviour is proposed in this work. The results call thus for further studies to determine the actual advantage introduced by mobility in real-world scenarios.
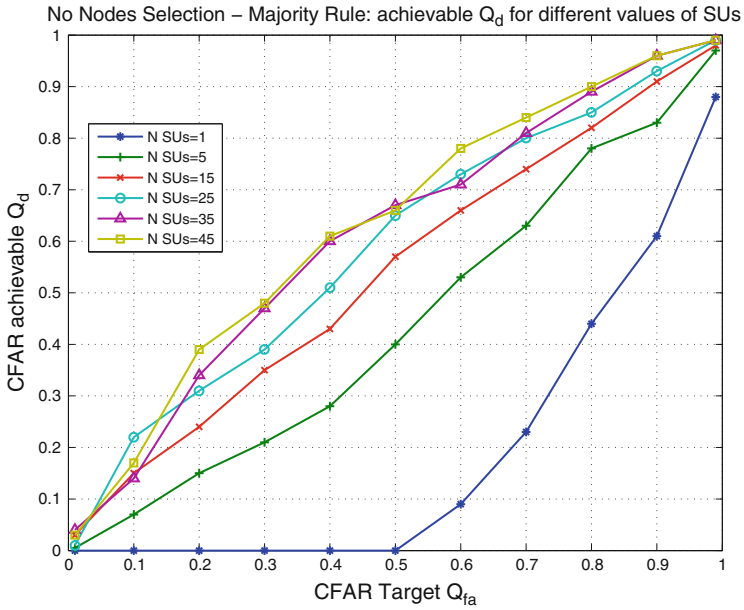
Fig. 11.12 ROC for CSS + Majority with 'No Nodes Selection' scheme
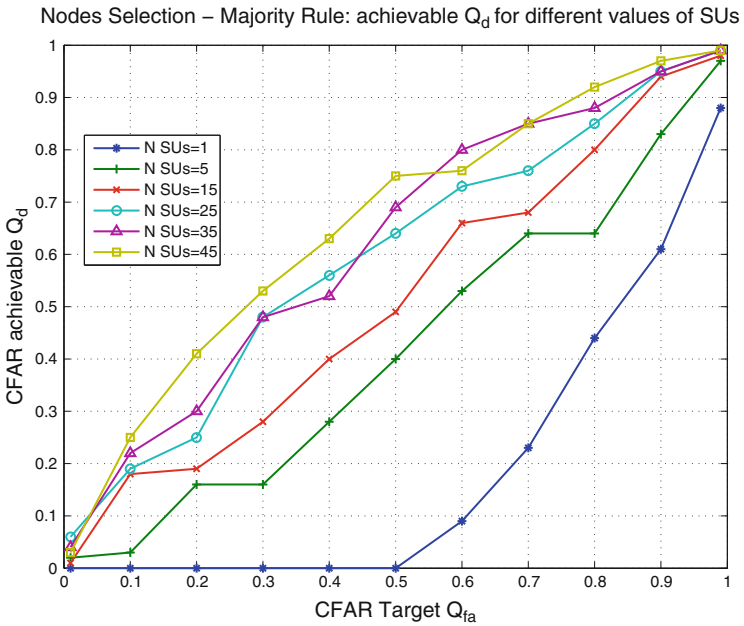
Fig. 11.13 ROC for CSS + Majority with proposed 'Nodes Selection' scheme
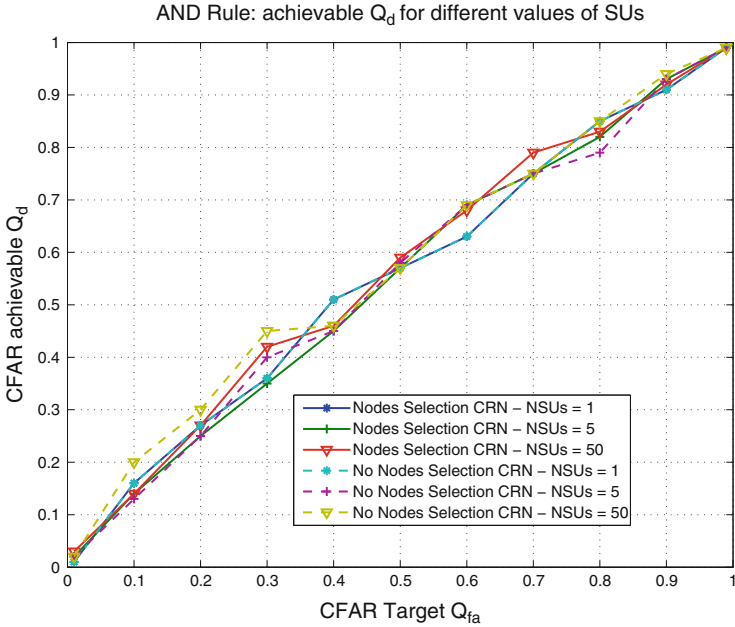
Fig. 11.14 ROC for CSS $+$ AND for proposed 'Node Selection' scheme vs. 'No Node Selection' scheme
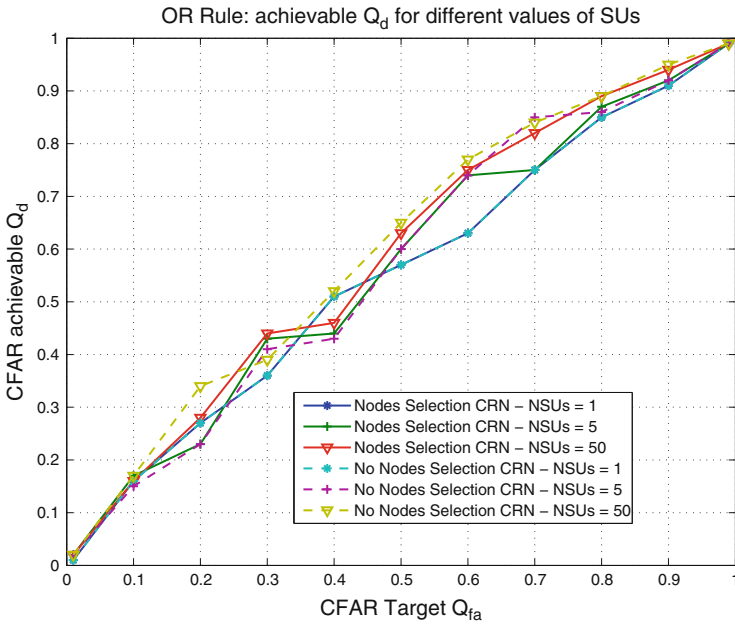
Fig. 11.15 ROC for CSS $+$ OR for proposed 'Node Selection' scheme vs. 'No Node Selection' scheme
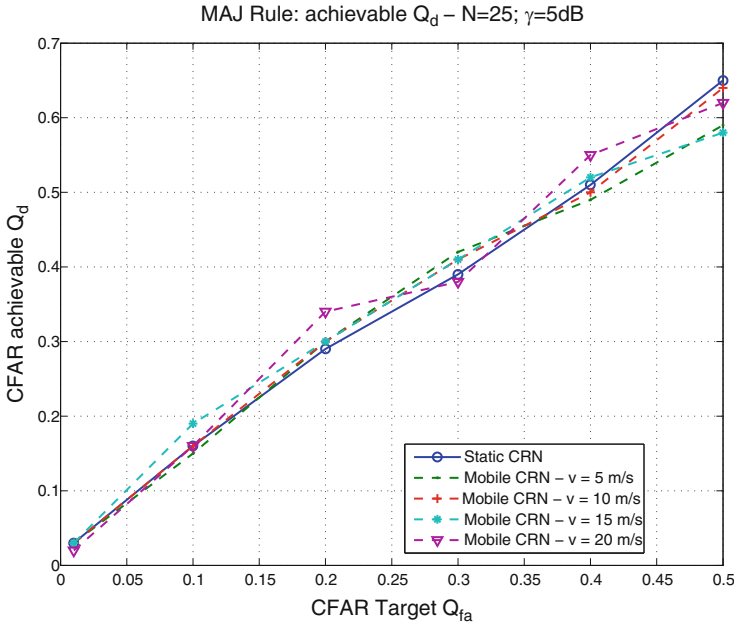
Fig. 11.16 CSS performance comparison for static and mobile CRNs

11.6 CR Simulator: Availability and Possible Collaborations

This chapter presents an ongoing effort towards the development of a CR simulation framework, to be used in the design and the evaluation of protocols and algorithms. The implementation of new features and functionalities of the simulator is still an ongoing activity at the ACTS Lab, within the DIET Department at the School of Engineering of the Sapienza University, in Rome. For this reason, collaborations and joint research efforts are highly desirable and welcome.

References

1. Axell, E., Leus, G., Larsson, E.G., Poor, H.V.: Spectrum sensing for cognitive radio: state-of-the-art and recent advances. IEEE Signal Process. Mag. **29**(3), 101–116 (2012)
2. Ghasemi, A., Sousa, E.S.: Collaborative spectrum sensing for opportunistic access in fading environments. In: First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), Baltimore, pp. 131–136, Nov 2005

3. OMNeT++ Website: http://www.omnetpp.org/
4. Koepke, A., Swigulski, M., Wessel, K., Willkomm, D., Klein Haneveld, P.T., Parker, T.E.V., Visser, O.W., Lichte, H.S., Valentin, S.: Simulating wireless and mobile networks in OMNeT++: the MiXiM vision. In: 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SIMUTOOLS'08), Marseille, Mar 2008
5. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. IEEE **LV**, 523–531 (1967)
6. Caso, G., De Nardis, L., Ferrante, G.C., Di Benedetto, M.-G.: Cooperative spectrum sensing based on majority decision under CFAR and CDR constraints. In: 25th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'13), London, Sept 2013
7. MiXim Website: http://mixim.sourceforge.net/
8. Wessel, K., Swigulski, M., Koepke, A., Willkomm, D.: MiXiM - the physical layer, an architecture overview. In: 2nd International Workshop on OMNeT++ Rome, Mar 2009
9. Caso, G., De Nardis, L., Holland, O., Di Benedetto, M.-G.: Impact of spatio-temporal correlation in cooperative spectrum sensing for mobile cognitive radio networks. In: 10th International Symposium on Wireless Communication Systems (ISWCS'13), Ilmenau, Aug 2013
10. Moran, P.A.P.: Notes on continuous stochastic phenomena. Biometrika **37**(1), 17–23 (1950)
11. Min, A.W., Shin, K.G.: Impact of mobility on spectrum sensing in cognitive radio networks. In: Proceedings of the 2009 ACM workshop on cognitive radio networks (CoRoNet '09), New York, pp. 13–18 (2009)
12. Arshad, K., Moessner, K.: Mobility driven energy detection based spectrum sensing framework of a cognitive radio. In: Proceedings of UKIWCS 2010, Delhi, Dec 2010

Chapter 12
Designing a CR Test Bed: Practical Issues

**Andrea Fabio Cattoni, Jakob Lindholm Buthler, Oscar Tonelli,
Luiz A. da Silva, João Paulo Miranda, Paul Sutton, Floriana L. Crespi,
Sergio Benco, Alberto Perotti, and Daniel Riviello**

Abstract Research on intelligent and reconfigurable wireless systems is in continuous evolution. Nevertheless, in order to fix some keystones, more and more researchers are entering the idea of research-oriented test beds. Unfortunately, it is very difficult for a wide number of research groups to start with their own set up, since the potential costs and efforts could not pay back in term of expected research results. Software Defined Radio solutions offer an easy way to communication researchers for the development of customized research test beds. While several hardware products are commercially available, the software is most of the times open source and ready to use for third party users. Even though the software solution developers claim complete easiness in the development of custom applications, in reality there are a number of practical hardware and software issues that research groups need to face, before they are up and running in generating results. With this chapter we would like to provide a tutorial guide, based on direct experience, on how to enter in the world of test bed-based research, providing both insight on the issues encountered in every day development, and practical solutions. Finally, an overview on common research-oriented software products for SDR development, namely GNU Radio, Iris, and ASGARD, will be provided, including how to practically start the software development of simple applications. Finally, best practices and

A.F. Cattoni (✉) • J.L. Buthler • O. Tonelli
Aalborg University, Frederik Bajers vej 7, 9220 Aalborg, Denmark
e-mail: afc@es.aau.dk; jlb@es.aau.dk; ot@es.aau.dk

L.A. da Silva • J.P. Miranda • P. Sutton
Trinity College Dublin, Dunlop Oriel House, Dublin 2, Ireland
e-mail: dasilval@tcd.ie; cruzlopj@tcd.ie; suttonpd@tcd.ie

F.L. Crespi • S. Benco • A. Perotti • D. Riviello
CSP – ICT innovation, Turin, Italy
e-mail: floriana.crespi@csp.it; sergio.benco@csp.it; alberto.perotti@csp.it; daniel.riviello@csp.it

examples of all the software platforms will be provided, giving inspiration to researchers on how to possibly build their own customized systems.

12.1 Introduction

The Cognitive Radio (CR) concept has been around for about 13 years now [38]. As any other concept, it was expected to pass through a theoretical phase first and then move towards practical developments. Nevertheless, the complexity and broadness of the involved topics slowed down this process, making CR experimental test beds a rare beast. On the other side, industrially driven wireless standards are quickly incorporating intelligence and automation in their specs, at the point that one could claim that Long Term Evolution–Advanced (LTE–A) is a CR standard.

All the above mentioned issues pushed individual researchers and groups to move into experimental research, for trying out, in real life, the effectiveness of their algorithms and techniques. Software Defined Radio (SDR) is currently regarded as the main technology enabler supporting the implementation of such experimental radio prototypes. The market is nowadays offering a vast variety of research platforms [14, 41, 42, 48], ranging from fairly low to very high prices, involving of course a change in performance as well. Together with what can already be considered Commercial-Off-The-Shelf (COTS) hardware solutions there are also associated several software (SW) ones. SW platforms for research purposes are generally open source, and freely available. But even in the case of SW, there are several differences in the available products, that do not make them fit all the different research needs.

For a newbie in the SDR and experimental research world it is very difficult to find a proper direction for developing his or her own test bed or application. One of the first issues that a researcher could encounter is, for example, finding the hardware (HW) platform that fits the need, without overshooting in complexity and price, nor undershooting and ending up with a platform that does not fit the requirements.

This was just a very simple example of all the practical problems that have to be faced every day by CR experimental researchers. Furthermore, the scientific aspect requires to be carefully analyzed, in order to provide solidity and consistency to the results generated with the test bed. Several aspects like the control of the experimental environment or the repeatability of the experiments and the comparability of the obtained results can create non trivial problems when multiple partners and multiple test beds are for example involved in a project.

The goal of this chapter is to guide researchers into the world of the experimental research, providing a tutorial guide on scientific and development issues encountered in years of practice by the authors. The hope is that more researchers could joint the experimental community, creating a vast network of test beds that could move faster thanks to shared experience.

12.2   Evolution of Test Bed Research

In this Section we would like to introduce an overview of the existing experimental researches and platforms present in both the literature and the market. Furthermore, we would like to present a point of view on how the experimental research has evolved in the past years, and where it is currently heading to. In the end, we would like to present the scientific needs of test bed based activities, in order to help the reader to understand which issues a research group potentially interested in starting development needs to consider for designing the experiments.

12.2.1   Trends in Implementation-Based CR Research

Research on CR has produced a number of applications that take advantage of the radio's capability to react intelligently to its current operating environment. Notwithstanding the value of these contributions towards a better understanding of CR, most of them rely on theoretical and/or simulation work. Simulation is the performance evaluation tool most widely used in this case, and research groups often build custom simulators to demonstrate the soundness of their own contributions. However, to be trustworthy, simulation experiments need to be repeatable, unbiased, rigorous, and statistically sound. Regrettably, the reliability of simulation studies has not been improving [44] and has even further deteriorated in some cases [34].

Implementation and emulation can be used as alternatives to pure simulation. Implementation means that the experiment runs a fully functioning protocol stack on top of a real-world platform, while emulation is the term used to denote experiments resulting from a combination of simulation and implementation [2]. Implementation adds value to theoretical modelling and analysis, as it can be employed to validate the design of algorithms, protocols, software, and hardware under a genuine radio frequency (RF) environment. This approach may be rewarding in the following ways:

- Results obtained through implementation are more tangible than those limited to theoretical and simulation-based research. Implementation-based research is therefore particularly valuable to research collaborations with industry and to the standardization process.
- The future of CR strongly depends on regulatory bodies. Implementation activities have proven successful in creating dialogue oportunities with regulators on spectrum licensing issues.
- Real-world insights gained through implementation-based experiments can be fed back into the theory and into simulation models.

In recognition to the above advantages, implementation in CR research has been ramping up recently. Meaningful insights have been gained through experimentally-driven research but these, especially in the context of CR networking, are not

Table 12.1 Evolution of CR test bed research – some issues & trends on orthogonal frequency division multiplex (*OFDM*), dynamic spectrum access (*DSA*), artificial intelligence (*AI*), cooperative spectrum sensing (*CSS*), and geo-location/database access (*GDA*)

Subject	Limitations	Trends
OFDM	Accounts for the majority of waveforms used The need for highly linear power amplifiers increases the cost of user terminals	Research on SC waveform design and SC-FDMA will further evolve pushed by 3GPP LTE
DSA	Most platforms are designed free of license Analysis in the absence of real-world incumbents does not reflect future operation conditions	Experimenters will tend to avail of trial licensing options as these allow experiments in licensed spectrum
AI	Demos showcasing the use of AI features in spectrum selection are entirely lacking Field seems to be not mature enough	Experiments will involve networks, distributed resources, and enforce collaboration among nodes
CSS	Most demos rely on semi-blind techniques Centralized approaches are the most popular Single-digit numbers of nodes widely adopted	test beds will be expanded to consider distributed, large-scale, multinode scenarios
GDA	Better suitable for spatially static scenarios Use in combination with sensing can improve performance but may imply higher costs	Real-world trials will be carried out to determine the costs associated with such hybrid solutions

without their limitations. Trends on how implementation will evolve in the coming years can be drawn on the basis of such limitations [43]. Table 12.1 summarizes some of these limitations along with their corresponding trends.

Waveform design based on OFDM remains a subject of interest but research on single carrier frequency division multiple access (SC-FDMA) will further evolve, especially after the adoption of the latter as multiple access scheme in 3GPP Long Term Evolution (LTE). DSA will also continue to be a hot topic, with experiments in licensed bands soon becoming mature enough to provide a deeper understanding about coexistence among DSA systems and real-world incumbents.

Most entries of Table 12.1 relate to CR networking. Here the first limitation has to do with the lack of experiments offering AI features. Such features lend themselves better to experiments involving resource allocation and enforcing collaboration among nodes. A second limitation is due to the need to determine the feasibility of hybrid coexistence solutions combining GDA with spectrum sensing. Single-node sensing suffers from several issues well understood today, so the sensing part of hybrid solutions will likely rely on approaches that – like CSS – obtain spatial diversity. While only 10–20 nodes can realize the full benefits of CSS [36], the number of nodes currently considered in networking experiments is limited to two or three. This is typically not enough to uncover most networking issues. In the future, experiments will consider distributed, large-scale scenarios where up to a few hundreds of nodes are deployed over a wide geographical area.

The current trend of moving experimentally-driven research on CR from the PHY layer to more network-oriented applications can be illustrated in terms of the Future Internet Research and Experimentation (FIRE) initiative [17]. FIRE aims to promote experimental research on networking concepts and architectures for the future Internet and, in parallel to that, build large-scale facilities to support experimentation in the medium and long term. Basically, the idea is to federate existing and new test beds allowing experiments on future Internet technologies. Such a one-of-a-kind infrastructure will be made available without placing on experimenters the high costs, complex logistics needs, and other burdens associated with real-world test bed implementation.

Among the FIRE projects, the Cognitive Radio Experimentation World (CREW) project is particularly relevant as it will lend lessons learned and solutions developed within its scope to illustrate the discussion in Sect. 12.2.3. The main target of CREW has been to federate five test beds, each offering different features in terms of operation bands, wireless standards, and platforms (both hardware and software) [10]. The CREW consortium, composed of eight partners from academy and industry, designed the federation so that the combination of individual test bed features makes several practical scenarios readily available to experimenters.

12.2.2 Review of the State of the Art

Test beds have been widely used in experimental research to validate new concepts and ideas stemming from theoretical research in wireless communications. Due to the broadness of research in this field, different test beds have been developed focusing on distinct aspects of wireless communications: the physical layer, where broadband transmission and MIMO systems prevail, the MAC layer and multiple access techniques, advanced radio resource management techniques and cognitive radio.

In the context of physical layer research, topics like multiuser MIMO, inter-cell cooperation and interference management are today very active. Many experimental test beds have been reported in the literature, e.g., [3, 8, 9, 16, 26, 30, 35, 40, 46, 57]. In most cases, these test beds consist of a single base station equipped with multiple antennas which makes possible to exploit the spatial diversity in order to increase the spectral efficiency. Few of them include multiple base stations [25, 35]. In these cases, the advantages of inter-cell coordination and interference management techniques (e.g, interference alignment and cancellation) can be shown to lead to major improvements in spectral efficiency.

Cognitive radio is the other field where most experimental research activities are being carried out through test beds [4, 27, 37, 53, 56, 58]. Here, the main focus is the dynamic use of spectrum by networks of systems coordinated by cognitive entities which reside on the network nodes. Primary users must be accounted for in the test bed by allocating resources for their emulation. Moreover, spectrum sensing is needed for the detection of primary users. Finally, the cognitive layer,

consisting of the several functional elements that enable enhanced adaptation through learning, have to be accounted for. Each of these issues has been faced and different solutions have been proposed, yielding results with different, often complementary, characteristics. Starting from the available test beds, the Cognitive Radio Experimentation World (CREW) test bed [54] has been developed by federating five test beds available at different locations throughout Europe. This project will be presented in detail in the following of this chapter. Several of the aforementioned activities are being developed using a number of hardware radio front-ends and of software radio platforms and tools, which are available for free in some cases.

Among the most widely employed hardware front-ends, the Universal Software Radio Peripheral (USRP) [14] has to be mentioned. It features a digital baseband/IF section and a complete set of analog RF modules covering the whole range of frequencies from baseband to around 6 GHz. Its mate software toolkit is the well-known GNURadio [55].

In systems developed using the GNU Radio toolkit, the digital baseband is fully implemented in software. Typically, the software-defined baseband delivers the time-discrete complex envelope of the modulated signal to the front-end, which only performs the necessary interpolation/decimation and filtering needed to reach the D/A or A/D converters operating at some intermediate frequency (IF). The final frequency conversion to/from RF and amplification are performed by the analog RF modules.

Although widely used, the USRP lacks some essential features for the implementation of full-duplex real-time transceivers, like accurate synchronization needed in MIMO systems,[1] and a low latency in the communication between the SDR process running in the workstation and the front-end, which makes very hard to implement transceiver with very tight timing constraints (e.g., TDMA and CSMA transceivers) without modifying the devices' hardware.

Several projects have been developed to overcome these limitations. Most of these address multi-antenna systems by accommodating several RF modules into a single front-end, like the WARP software radio platform [47], or reduce the latency by using parallel bus connections (e.g., PCIe) between the workstation and the front-end [42, 48]. Others conform to a processing model wherein the baseband signal processing algorithms are executed by on-board processing resources very close to the front-end unit [41].

As for the proposed signal processing model, some of these platforms fully rely on software, leaving to hardware (typically, a FPGA) only the basic interpolation, decimation and high-rate filtering operations needed to cope with the high A/D and D/A sampling rates. In other cases, powerful FPGAs or DSPs are made available

---

[1]Although some of the USRP products feature characteristics that enable synchronization across multiple devices to perform MIMO processing, it becomes somehow expensive and cumbersome to build MIMO systems starting from several of these devices.

to perform the most computationally intensive tasks relieving the CPUs from heavy processing loads [41].

12.2.3   *Experimentation Methodology and Benchmarking*

The need for a common methodology supporting experimentation under test conditions constitutes a big issue in experimentally-driven research. The goal is to define a set of procedures to ensure that the output of experiments can be made comparable, repeatable, and controllable. The adoption of benchmarking frameworks allows experimenters to establish fair comparisons among different systems under test (SUTs). Comparisons of this kind call for a benchmarking framework that is standardized or, at the very least, widely accepted. No such a framework seems to be available in the CR test bed literature at the moment.

In the sequel we provide a set of guidelines that may be helpful when designing a benchmarking framework for CR test beds. To begin with, we put in place the terminology that will be used throughout the remainder of this section. We then discuss the requirements considered as key in the design of any benchmarking framework. Extensions specific to wireless test beds are also addressed. We wrap up the section with a description of an exemplary benchmarking framework created to fulfill such requirements.

12.2.3.1   Benchmarking Terminology

Wireless benchmarking can be defined as the measurement and evaluation of wireless networks, protocols, devices, or more general SUTs relative to a reference evaluation [19]. In the discussion that follows we examine the elements of that definition and introduce some benchmarking terminology.

The first benchmarking element is *measurements*. In the context of spectrum sensing, a measurement accounts for data collected by sensing agents so that a test statistic can be constructed at the end of each sampling period. Examples of sensing data may include energy, covariance matrix, estimates of power spectrum density (PSD), and raw samples.

The second element of benchmarking is the *evaluation* of the SUT. This calls for performance metrics that are classified as primary or secondary in accordance with their importance in the benchmark. Throughput, reliability, and sensitivity of the SUT are examples of primary metrics. Secondary metrics are different in that they can impact a primary metric but this impact cannot be directly observed by the experimenter. This is the case with metrics related to the wireless environment, e.g. interference, channel occupancy, etc. The collection of metrics used in the comparison determines the benchmark *score*. The translation of metrics into a benchmark score is also dependent of a set of benchmark *criteria*, which defines the focus and characteristics to be examined in the benchmark.

A benchmark framework is built from fundamental building blocks called *components*. In general, components can be grouped into the following categories:

- Generic components: Scenarios, criteria, parameters, metrics, and capability control are examples of generic components. They are meant to be reusable, whether on a standalone test bed or in individual test beds of a federation.
- Adaptors: They refer to resource descriptors, common result format, benchmarking control, and other interfacing components. Adaptors account for the communication vehicle between generic and test bed-specific components.
- Specific components: Such components are present in most test beds. They possess capabilities and/or require interfaces that are test bed-specific, e.g. resource manager or data storage.
- Virtual components[2]: These result from a combination of test bed-specific components and as such can provide a test bed federation with new capabilities on the basis of existing ones.

Having these terms in place, we can now define a *scenario* as a full description of a benchmark. A typical scenario consists of a set of performance metrics, evaluation criteria, along with all parameters and traces required to run an experiment.

12.2.3.2   Benchmarking Requirements

We now discuss a set of common requirements of a benchmark, in the context of wireless networks.

Comparability, Repeatability and Reproducibility: *Comparability* and *repeatability* account for the two most fundamental requirements of any benchmarking: The former ensures that two benchmarks carried out independently can be compared to each other, while the latter establishes the limits within which results obtained via repeated benchmarks can be considered equal. An extension of repeatability is *reproducibility*, which requires a benchmark to be repeatable not only on a single test bed but also reproducible on any test bed having the same capabilities. The requirements discussed so far are basic but pose additional burdens on any benchmarking solution for wireless test beds. Repeatability, for instance, is not trivial to guarantee in wireless environments due to uncertainties like type and color of the noise, multipath fading, and noise power. Reproducibility requires additional environment and test bed modelling.

Automated Performance Evaluation: The first nice-to-have extension that helps setting up and executing benchmarks is automated performance evaluation. This allows multiple experiments to be scheduled and assessed in an automated fashion, with the parameter space optimized automatically [21].

---

[2]Along with its supporting interfaces the mix & match approach for virtual components is one of the fundamental contributions of CREW. See [20] for a more detailed discussion on the subject.

Environment Monitoring and *Environment Modelling:* The assumption of an experiment is running free of non-SUT interference may be too strong to be practical in wireless environments. In this case, a validity check can be carried out to identify potential activity of devices not part of the SUT. Spectrum sensing can be used as a means to determine such external influences, helping experimenters in the task of achieving acceptable levels of comparability and repeatability.

A methodology for validity check is proposed in [19]. The methodology, referred to as environment monitoring, consists of detecting potential non-SUT interference in three phases. The first phase relies on blind signal processing techniques to determine the channel occupancy *before* the experiment is run. Upon detecting activity in the channels of interest, the system can be set to postpone the experiment until the channel is deemed to be idle or switch to another idle channel. The second phase takes place *while* the experiment is running. The ability to distinguish among SUT and non-SUT signals may be required in this case, so the validity check needs to rely on signal-specific techniques. The third phase is carried out *after* the experiment is terminated. The rationale motivating for this last check is that any non-SUT activity identified at this point was likely also present during the second phase.

Common Data Collection: A methodology for collecting experimental data constitutes another crucial extension in the benchmarking process, whereby input/output data relevant to the experiment can be defined using a common data structure. If most of the relevant data relates to spectrum sensing, a natural way to go is to rely on the IEEE 1900.6 standard [1]. The standard defines a logical interface along with supporting data structures for the exchange of data among wireless sensors and their clients. The abstract nature of IEEE 1900.6 is particularly appealing as it allows sensor (and sensing algorithms), client, and the communication link between sensor and client to be designed on an unconstrained fashion. Specifically, the data relevant to sensing-related experiments can be grouped into the following categories [20]:

- Abstract: This is a structured description of the experiment in terms of what was performed, how it was performed, and who performed it. The abstract is valid for the entire experiment set.
- Meta-information: This data describes all parameters used in the experiment in compliance with the IEEE 1900.6 standard, e.g. hardware, software, application, traces, and processing tools.
- Iterations: Start/end of timestamp, values assigned to the parameters in the meta-information, files, are all examples of iterations used to provide additional information about traces.

Common Storage: Whenever experiments are run on (and benchmarked over) multiple test beds, the creation of a federation database for storage may also have its advantages. The idea is to facilitate the use of output data acquired in one test bed as input data in another test bed. Once populated the repository can be made publicly

available, so external experimenters working in the field of CR and CR networks can be given the chance to benefit from its content [11, 23]. The content, typically a collection of different data types, may vary depending on the research objectives behind the repository. The repository available through [22], for instance, contains full experiment descriptions, traces, wireless background environments, processing scripts, performance metrics, and bechmarking scores.

12.2.3.3 Exemplary Benchmarking Framework

Once the benchmarking requirements are defined, the next task at hand is the design of a framework that fulfills them. In recognition to the importance of benchmarking requirements and specific federation needs, it is advisable that the benchmark process be based mostly on generic components, with hardware and test bed-specific tools reduced to a minimum. In general, the workflow of such a process begins with an input accompanied by a benchmark definition. A score is computed on the basis of the output produced after the benchmark execution, which in turn should be the only step controlled by test bed-specific resource management tools. Input and output should be both compliant to a common data format, so the exchange of configuration, metrics, and results among project partners can be made easy.

Figure 12.1 illustrates the workflow above. A *usage scenario* is motivated by general research objectives, while a *test case* addresses a particular aspect of a given usage scenario. Examples of usage scenarios may include context awareness for CR networking, cooperation among heterogeneous networks in licensed bands, or CR systems incorporated into cellular networks [18]. Next, and still on the experiment level, all descriptive elements associated with the selected test case are written in a common data format. The full description obtained thereby allows to the automated derivation of scenarios. These drive the whole benchmarking process in that they set up generic components such as parameter space, evaluation criteria, and, eventually, the required traces to be used as input data.

The *benchmark control* adaptor defines different benchmarks on the basis of the subset of parameters and traces made available at the output of the capability control component. This adaptor also triggers the execution of the different benchmarks, which can be carried out through the *resource manager* or using any other management interface. In either case the control methods used are defined in the *resource descriptors*. Experimental results are first stored in a dependent *datastore* from where they can be subsequently translated into a common data format. Benchmark criteria then determine the metrics to be applied in the analysis. Optionally, a *result feedback* link can be implemented on the parameters and traces used in the experiment. This option provides advanced features such as fully automated benchmarking. The process terminates after a benchmark score based on the selected metrics is generated.
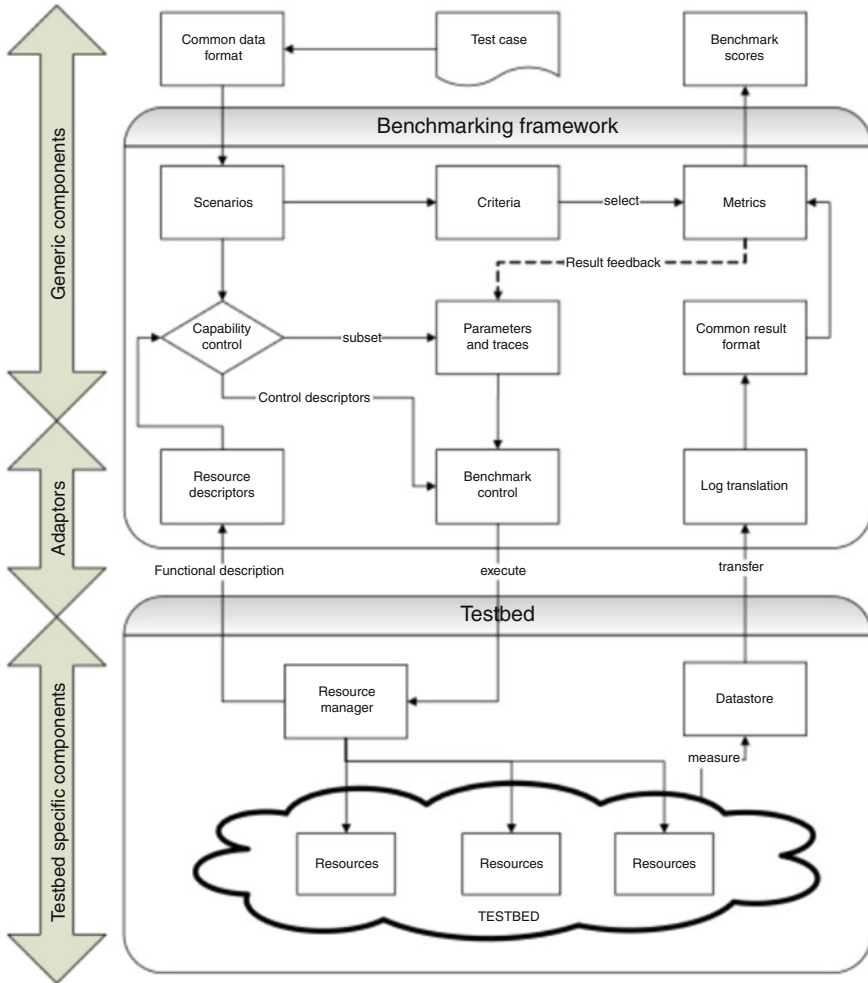
Fig. 12.1 An examplary benchmarking framework [19]

12.3   Practical Research Activities with a Test Bed

Many considerations regarding the requirements of the test bed and how to full fill them must be taken before jumping into experimenting using test beds. This section will raise the key issues to be addressed such as understanding the requirements of your own test bed and use these to choose the right hardware solution. Even though the hardware fits the test beds requirements imprecisions in the hardware will cause the need for some meddling with the hardware before the experimentation process can begin.

12.3.1 Understanding the Requirements of Your Own Research Application

The first step towards the implementation of a SDR-based test bed is the understanding of the general requirements to be satisfied by either the hardware and the software platform, since different types of applications may lead to totally different requirements. As an example, note that test beds offering a full protocol stack will deeply differ from those that provide only physical and link layer implementations. The latter are often aimed at research on cross-layer approaches, in which the protocol stack may be totally customized around a set of specific constraints.

In an attempt to list, in order of importance, different application requirements for a test bed, we start from investigating the number and type of nodes involved. This constitutes a rationale for the management of available resources. In fact, given a fixed budget, the requirement to emulate a densely populated scenario could result in less processing resources and a limited number of interfaces available at each node. Otherwise in the simplest case of a point-to-point system with two nodes, each of these could feature enhanced capabilities. Moreover, different types of nodes could be involved, as in the case of a single- or multi-cell scenario with mobile user terminals. In such case, the requirements may differ significantly from node to node: base stations (BS) should be capable of supporting multiple users simultaneously, of coordinating with other BS through suitable interfaces, of multi-antenna broadband transmission. Terminal emulators should be small and lightweight to allow mobility, yet powerful enough to support the heavy processing load of multiuser MIMO reception and channel decoding.

Each node typically features one or more interfaces with different characteristics: wired or wireless connectivity, maximum bit rate, standardized or proprietary connectors and protocols. Common examples of wired interfaces are Ethernet, USB and other serial interfaces, while for the wireless case we have 802.11a/b/g/n/ac, 802.15.1 (i.e., Bluetooth), 802.15.4 (a.k.a. ZigBee), only to mention a few. Another important type of interface is represented by the PCI and PCIe parallel bus interfaces. These interfaces reach maximum bit rates of several Gbits/s, while featuring a very low latency: few microseconds, typically. The chosen type of interface determines the latency, maximum bit rate to/from the workstation, enables MIMO capabilities through distributed synchronization across multiple front-ends.

Another set of considerations concerns mobility requirements: in the case of networks characterized by high mobility, the energy consumption of each node is crucial. Self-powered nodes often imply a limited availability of computing resources that may limit the flexibility of the overall system, although it is not uncommon to find powerful and energy-efficient multi-core embedded processors even in many battery-powered devices of daily use like smartphones, tabs, etc. These devices can easily be turned into signal processing engines for SDR test beds, if equipped with suitable interfaces.

In the following list, we highlight some of the most relevant aspects to be considered when defining test bed requirements:

- Node types: nodes are all of the same type/are of different types
- Node density: network of nodes, point-to-point system.
- Network topology: mesh, star, tree, . . .
- Architecture: centralized, distributed, cellular (single or multiple cells, inter-cell coordination)
- Mobile or static scenario
- Single or multi-antenna/multi-sensor node
- Processing capabilities available at each node
- Scalability, with respect to all previous features

Possible solutions to such practical issues will be addressed in the following part of this chapter.

12.3.2 Choice of the Hardware

When designing an SDR test bed several factors must be taken into account: the choice of the hardware is probably one of the first and most important decisions to be made. The price and capabilities of the hardware devices available on the market may vary considerably. Specific solutions should be selected in relation to the characteristics of the desired test bed setup. Test bed configurations focusing on network applications for example, employ a large amount of nodes which typically have different hardware requirements (in terms of processing capabilities, components accuracy, RF support) in comparison to e.g. a MIMO test bed. In the first case the price per node becomes much more critical than in the second. Another consideration relates to the required skills for implementing a test bed with a specific hardware setup. DSP and FPGA programming are typically quite different skills to possess from an higher-level object-oriented programming. In this section we will try to analyze all aspects related to the choice of the hardware platform which may impact the development of an SDR test bed.

The *radio-frequency (RF) capabilities* of the hardware should be identified by understanding the target application for the test bed. The choice of the supported operational bandwidths is for example a critical issue: it is indeed rather intuitive to understand that the available RF front-ends completely determines which frequency bands the test bed can operate in. Some hardware solutions aim at greater flexibility supporting a large range of frequencies, others provide more optimized support in specific operational bands. Always taking into consideration the cost per device, in most of the cases the choice is about compromising the spectrum coverage with filtering and sampling precision. Unless referring to a specific technology, such as WiFi, it might be useful to select a hardware supporting frequency bands suited for experimentation purposes, thus with minimal external interferers. In this case it is important to remember that some bands have restricted usage by law and this

varies from country to country. If the test bed requirements foresee the support of MIMO configurations, the possibility of having multiple RF front ends is a rather fundamental attribute for the hardware. MIMO applications also require means for clock synchronization and precise timing control of samples from/to multiple front ends. Applications featuring time division duplexing (TDD) typically require the hardware to manage the stream of timestamped samples. Frequency division duplexing (FDD) on the other hand, requires multiple independent oscillators such that multiple different carrier frequencies can be used parallel.

In order to identify the required capabilities from the RF front-end it is important to understand the characteristics of the signals and waveforms we intend to use with the test bed. For example OFDM transmissions and other waveforms with high Peak-to-Average-Power-Ratio (PAPR) have typically higher requirements in terms of quality of the power amplifier and the IQ balance.

The *connection between the RF front-end and the processing unit* is also an extremely important element to evaluate. The turnaround time in sending and receiving data to/from the hardware, and the maximum achievable data rate may vary considerably depending on interfaces utilized. In general terms the streaming of samples to and from an host PC is less performing in comparison to managing the data processing directly on an integrate hardware solution. Specific radio applications (e.g. CSMA) may be sensitive to these aspects, requiring fast-response in order to quickly access to the medium. In these cases an embedded solution (e.g. ARM processor connected closely to the data acquisition unit) is preferrable providing reduced latency. Alternatively a soft-core processor implemented directly on an FPGA can also provide similar support. In contrast if latency and throughput are not major issues PC-based hardware configurations may provide greater advantages in terms of design flexibility and fast prototyping. The usage of widely common interfaces such as USB and Ethernet favors the usage of commercial-off-the-shelf hardware which typically ensures better compatibility, reliability and support in time in respect to more customized hardware solutions.

The choice of hardware components is just one of the steps necessary in the development of a test bed for research purposes. However, given the amount of economical and human resources typically involved in this kind of research projects, it is fundamental to act with a clear and long-term vision in mind. Replacing inadequate hardware may be extremely costly; scalability must be taken into account since the beginning, required skills and competences to manage the hardware are not of secondary importance also considering that several researchers may follow in its development and usage.

12.3.3 Practical Test Bed Issues

Anyone with a minimal experience in experimental research can tell about the amount and the variety of practical issues suddenly appearing when trying to implement theoretical concepts in real-world test beds and prototypes. Research

with cognitive radios is clearly no exception. Managing radio test beds with multiple devices and advanced hardware capabilities requires the definition of practical procedures enabling and efficient execution of the experimental trials. According to the scope of the research investigation different typologies of CR test beds exist; analysing their main characteristics may be helpful in understanding potential problems in the practical execution of the experiments thus inspiring prior planning.

A first and important aspect relates to the mobility of test bed terminals. In a *mobile test bed* nodes should typically be able to operate while moving or being transported. In practice this means that all the hardware equipment involved needs to be loaded over movable structures or mounted into vehicles. Two features are critical: agility and connectivity. In the first case it is important that test bed nodes are packed over a robust, agile and small-factor solution in order to be able to move the equipment easily and safely. Customized trolleys, wearable solutions and ad-hoc vehicles are typically the choice. While on the move it is also rarely possible to count on cables. Both power supplies and network backhaul should be then re-engineered by means of portable batteries and wireless configurations. Please carefully investigate the deployment environment before setting up a test bed. Particular attention should be given to the pavement conditions (if employing trolleys), dimensions (if wireless coverage is needed, it is important to ensure the backhaul to adequately cover the entire area of interest) and mobility aspects (if driving or running experiments in public places).

With *stationary test beds* many of the previous issues can be easily addressed, a number of recommendations can however be made in relation to the execution of experiments. Static experiment trials often foresee the occupation of facilities for relatively longer amount of time. The first practical issue is then to ensure the control of the experiment area for the entire duration of the experiments. Rooms reservation and access control are some examples. If nodes are not supposed to be moved, supplying power and connectivity via cables is a good option. The availability of power plugs and network access points may however be not sufficient. It is good practice to carefully estimate the need of power sources and connections and possibly bring multiple plugs, routers, switches in order to overcome the limited support of the deployment facilities. When dealing with cables, their inadequate length is also a frequent and overlooked problem.

Besides managing its deployment, the control of the test bed network is the second major issue in executing experimental trials. As mentioned before a backhaul connection is typically an essential part of the test bed. Relying on an IP based network with static addresses assigned to the host PCs might be a very convenient way to provide access and identification to the test bed machines. Establishing a Virtual Private Network (VPN) can also provide remote access to the test bed resources. The direct access to the test bed can be achieved through secure shell (SSH) connections. SSH clients are available for all major operating systems, thus allowing to connect to OS-specific test bed machines from a wide range of platforms. Care must be taken when controlling the test bed remotely (e.g. by directly handling multiple SSH sessions) to avoid experiment disruption due to a connection failure. A hybrid local/remote solution ensuring greater reliability can

be achieved by implementing a client/server architecture in the test bed. A single local machine in the test bed (the server) is responsible to deliver commands and acquire data. Remote control is granted through this machine thus providing a more robust solution to the remote host-server connection failures.

As the test bed develops and expands (testifying a fruitful experimental research), the maintenance of its elements and the management of its utilization among multiple researchers are likely to become more urgent tasks than what probably looked like in the beginnings. In this respect, one of the most common problems relates to ensuring up-to-date compatibility of the test bed applications with updated hardware and software components. The periodical verification of the operational status of software libraries and hardware firmwares is a good practice to implement. In order to schedule the availability of the test bed and related facilities, centralized booking systems and calendars are also recommendable.

Knowledge transfer is another aspect that in the practicalities of the test bed management is often overlooked. Quite a number of developers/researchers may need to operate with the test bed throughout time thus requiring a continuous process of training at the usage of the existing functionalities. It is good use to establish a Wiki document or other means (e.g. web pages, forums) in order to collect all the acquired prior experience with both the test bed, the experiments and maintenance procedures. Sensitive topics have proven to be:

- How to develop a new test bed feature
- How to access and process existing data
- How to handle software/hardware updates
- How to calibrate hardware
- How to initialize a brand new test bed node
- How to document the work

By addressing the practical issues of test bed management before the experimentation process is begun, it is possible to attack many problems before they arise. However, the practical issues are not the only to be addressed before experimenting.

12.3.4   Dealing with the HW

When using a set up consisting of a PC connected to an off the shelf front end, hardware imperfections and limitations might restrict the performance of the testbed. This section will describe some of these challenges and their impact. Where as Sect. 12.3.2 mentions some of the factors which cannot be altered by taking them into the test-bed design, this section will mention some of the issues which can.

Calibration of the RF hardware is often required when doing measurements since imperfections caused by e.g. the digital up converter, digital filters or the power amplifier, can cause the measured values to differ throughout the deployed nodes. This calibration can be taken care of in the software by correcting the values of the

received or transmitted signal based on a table of measured correction factors. An important factor is the hardware's center frequency, which can vary due to offset between the on-board oscillators in between the nodes. How much each node is offset can be found by transmitting single tones and then recording them on a high precision spectrum analyser. By noting the values in a table which the software can interpret, the values describing the offset can be used to correct the setting of the center frequency and hereby limiting the offset. Besides to the center frequency, the gain characteristics of the power amplifier throughout the supported band or in between nodes can differ. Similarly to the frequency offset, this will lead to e.g. incorrect spectrum sensing results. A simple Tx calibration procedure can be made by using an application which sweeps a single tone through the desired frequency band and a spectrum analyser which can cover the desired part of the spectrum and has a "hold" function. By letting the application sweeping through the spectrum and the spectrum analyser recording, it is possible to see the characteristics of the transmitter and hereby noting the offset in respect to the desired output. A calibration routine for Rx can be made by reverting the process.

Calibration can be a time consuming process if done manually and there is a chance that new drivers or firmware on the hardware can render the calibration factors useless. Therefore, it can be an advantage to make an automated calibration procedure. When creating an automated calibration procedure it is important to note that the precision of the calibration will be no better than the reference. Either an automated calibration procedure can be done by having spending time to calibrate a reference node very precise and then use this as a reference through a simple platform application. Another way of automating the procedure could be to use the GPIB [32] connection available on most high end measurement equipment to control the measurements and obtaining a high precision calibration. The advantage of the first is that it might be easier to synchronize the calibration procedure, whereas the latter is more precise.

Time synchronization between nodes can be difficult to achieve if the test bed is distributed over a large area. It is however required if the nodes have to do synchronous tasks such as switching frequency on a channel sounder application. One way to achieve this is by using the Network Time Protocol (NTP) where it is possible to achieve time synchronization down to the order of microseconds. Another, and more expensive option, is to use the Global Positioning Systems (GPS) time where a time precision can be achieved in the order of nano seconds.

Clock misalignment and *drift* are issues which should be addressed if the test bed are in the need of coordinated Tx/Rx. Using a common clock signal would be an optimal way to avoid this, however hard to implement in practice if the nodes are far apart. If the hardware either has a high precision clock or at least the option to connect an external clock input, the problem can be greatly mitigated. However, if there are requirements for high precision it might be worth the time to implement a distributed synchronization algorithm within the test bed itself.

12.4   SW Platforms

We would like here to concentrate on three open source platforms that can be considered the main enablers for CR applications development. These SWs, namely GNU Radio, Iris, and ASGARD, present several similarities in their design principles, though they are different in their target applications and goals.

All the mentioned frameworks provide a set of design tools, which allow managing the implementation of the processing tasks and the creation of a system architecture graph, by means of interconnected software elements. Although sharing common design principles, as the C++ implementation of processing components and the widespread use of XML descriptors for the definition of the system architecture, relevant differences exist especially in relation to the runtime execution and the data-passing characteristics. In Table 12.2 a quick comparison of the three chosen SWs is presented, in order to help the reader better understanding the differences.

12.4.1   Writing an Application

This section will describe how to design the communication system architecture and how to implement an executable application for experimental purposes using the previously described platforms; GnuRadio, Iris and Asgard. This includes description of how to implement a processing block, which means are available to connect and control the data flow, how to manage configuration parameters and available features to control the test bed.

12.4.1.1   Iris

Iris is a program and a set of libraries written in C++ which can be used to build software radios [13,52]. One of the major advantages of building a radio in software is the flexibility it provides. Iris is designed to support and exploit this flexibility or reconfigurability.

Iris is built using a plugin architecture. Each plugin is a library which does a specific job (e.g. data scrambling, OFDM modulation etc.) and which provides a generic API for the core Iris program to use. These libraries can be dynamically loaded at runtime and used within a software radio design. The main type of Iris plugin is called a *component*. Components typically process streams of data. They have input and output ports and work by reading data from one or more of their inputs and writing data to one or more outputs. In Iris, components run within an *engine*. The engine is responsible for loading the component library, initialising the component, providing input data to it, calling it to work on that input data, taking output data from it, destroying it and unloading the library when the radio

Table 12.2 Comparison of the available test bed frameworks

	IRIS	GNU Radio	ASGARD
Processing block design	– Divided into individual components with input and output ports. – I/O data ratios can change dynamically. – Component execution is managed by the internal Iris scheduler. – C++ implementation.	– Divided into individual blocks with input and output ports. – I/O data ratios need to be specified. – The block execution is managed by the internal GNURadio scheduler. – C++ implementation.	– Divided into individual components with input and output ports. – I/O connections defined as function pointers which can be connected to any type of object. – Do() method defines a single processing task. – C++ implementation.
Flow graph description	– XML configuration used to connect blocks in the flowgraph. – Iris framework manages loading and configuration of graph from XML descriptors. – Controllers exist outside of flowgraphs to support reconfiguration. – Graph loops are not allowed.	– Python scripts used to connect blocks in the flowgraph. – GNU Radio Companion is a GUI that enables a visual design. – Must have at least one source block (no input ports) and one sink block (no output ports). – Graph loops are not allowed.	– Described in the ASGARD Application using C++. – Modules creates an abstraction from threaded objects. – Settings loaded from XML file. – Components connected by the use of pointers to internal methods of the data containers.
Dataflow control	– *Engines* provide a modular abstraction for parallel threads. – *Phy Engines*: directional dataflow, sequential execution of components. – *Stack Engine*: bidirectional dataflow, parallel execution of components. – Iris *Events* support asynchronous control and data delivery to controllers.	– Handled by internal scheduler. – Each block in the flowgraph is executed as soon the input data is ready. – Centralized control: efficient dataflow, limited flexibility. – Designed for directional dataflows. – GNURadio messages enables asynchronous data delivery out of the flow.	– Multi threaded system handled by the OS scheduler. – Defined by which Components are loaded to which Modules. – CO's create an event-based control abstraction. – High flexibility at the cost of increased complexity.
Data-transfer	– Defined according to the Engine. – Phy engines use buffered data sets. – Stack engines support pointer-based data transfer.	– Input/output buffer copy (memcpy). – Data access managed by the framework. – Multiple dataset can be queued to a given block (multiple independent threads can run concurrently).	– Various data-transfer methods implemented through different CO's but can be used across different Components. – Thread safe data access. – Pointer based data transfer for high efficiency.

is shut down. Typically, a component in an Iris radio will run in a loop, repeatedly processing sets of data before it is destroyed and unloaded. Components can expose parameters which control how they operate and these parameters may be used to reconfigure a radio while it is running.

Iris currently has two types of engines (and thus two types of components) – the *Phy* engine and the *Stack* engine. Phy components typically operate on a stream of signal data which flows in one direction from input to output and execute only when called by their Phy engine. Examples include modulators and demodulators, channel coders and decoders, data scramblers etc. Stack components on the other hand may support bidirectional data, coming both from above and below. They run their own threads and can generate sets of data at any time, in addition to processing data coming from above and below. Stack component examples include complete MAC layers, network routing layers and data encryption layers.

Listing 12.1 Example XML configuration file

```xml
<softwareradio name="Radio1">
   <engine name="phyengine1" class="phyengine">

      <component name="filerawreader1" class="filerawreader">
        <parameter name="filename" value="testdata.txt"/>
        <parameter name="blocksize" value="140"/>
        <parameter name="datatype" value="uint8_t"/>
        <port name="output1" class="output"/>
     </component>

      <component name="ofdmmod1" class="ofdmmodulator">
        <port name="input1" class="input"/>
        <port name="output1" class="output"/>
     </component>

      <component name="signalscaler1" class="signalscaler">
        <parameter name="maximum" value="0.9"/>
        <port name="input1" class="input"/>
        <port name="output1" class="output"/>
     </component>

      <component name="usrptx1" class="usrptx">
        <parameter name="frequency" value="5010000000"/>
        <parameter name="rate" value="1000000"/>
        <parameter name="streaming" value="false"/>
        <port name="input1" class="input"/>
     </component>

   </engine>

   <link source="filerawreader1.output1" sink="ofdmmod1.input1" />
   <link source="ofdmmod1.output1" sink="signalscaler1.input1" />
   <link source="signalscaler1.output1" sink="usrptx1.input1" />

</softwareradio>
```
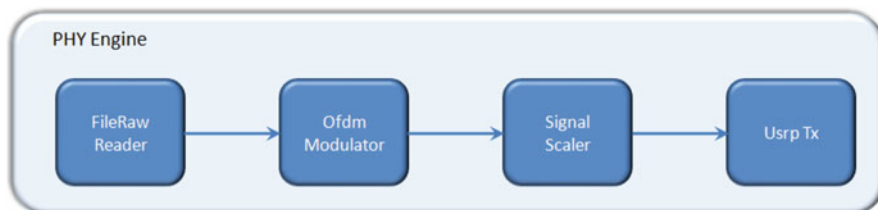
Fig. 12.2 Resulting radio configuration

In order to run a radio in Iris, an XML configuration file is used. This file tells the core Iris program which engines will be used to create the radio, and which components will run within those engines. It also includes the initial parameter settings for each component. An example XML configuration file for a simple OFDM transmitter is shown in Listing 12.1.This XML configuration specifies four Phy components which will run within a single Phy engine. In this radio, data is read from the file "testdata.txt", modulated into an OFDM signal, scaled in magnitude and transmitted with a specific carrier frequency and bandwidth by a radio front-end, in this case an Ettus Research Universal Software Radio Peripheral (USRP). The resulting radio is illustrated in Fig. 12.2.

You can see that some initial parameter values are provided for some of the components in the radio. If a component has parameters which are not specified in the XML configuration, these are set to default values (as is the case for the OFDM modulator here). An XML configuration file specifies the structure of an Iris radio when it is initially loaded and run. However, it can also be used to reconfigure a radio while it is running. This can be done for example by changing the value of a parameter in the file, saving it and prompting Iris to reload it. Iris will compare the configuration in the file with that of the running radio, find that one of the parameters has changed and reconfigure that parameter as required.

Using the XML configuration file, an Iris user can easily reconfigure a running radio. However, we often need a radio to reconfigure itself instead of relying on user input. This might be the case for example, when we wish to design a receiver which scans a set of channels for signals of interest. We have already discussed the main type of Iris plugin – the component. The second type of plugin which is used in the Iris architecture supports this self-configuring behaviour – the *controller*. Controllers are libraries which are loaded at runtime, just like components. However, a controller does not run in an engine and typically does not operate on streams of data. Instead, a controller has a global view of a running radio and can reconfigure any component in the radio at any time. For example, a simple controller could be used to scan frequencies in a wireless receiver as shown in Fig. 12.3.
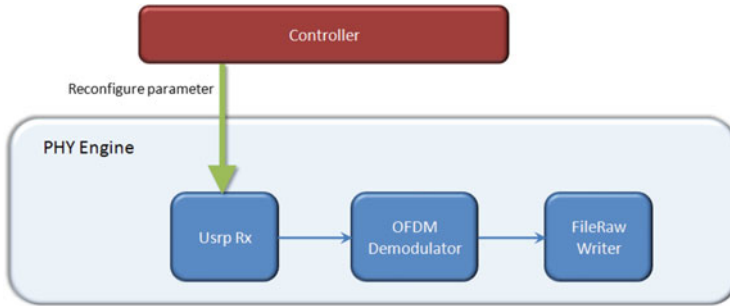
Fig. 12.3 Reconfiguring a radio using a controller
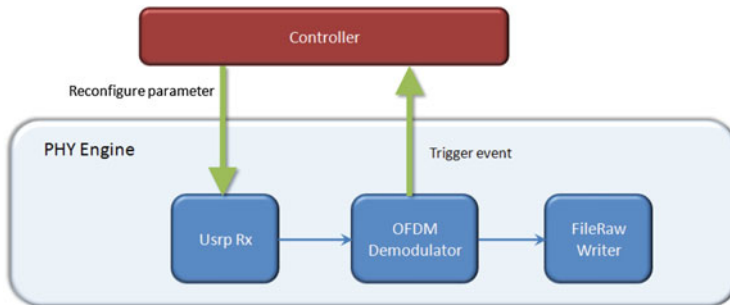
Fig. 12.4 Reconfiguring a radio in response to an event

Here, the radio consists of a Usrp front-end receiver, an OFDM demodulator and a file writer. The controller simply enters a loop, reconfiguring the receive frequency on the Usrp radio front-end and then sleeping for a set amount of time. Of course, with this design there is no mechanism for the radio to lock onto a received signal – it will simply continue scanning forever. Some mechanism is needed to allow a component to notify the controller to stop when a signal is received.

In Iris, this is what component events are used for. Building on the previous example, we could design the OFDM demodulator so that it triggers an event whenever an incoming signal is detected. Controllers can subscribe to events on specific components in order to be notified when that event is triggered. In this way, our controller gets notified when the OFDM demodulator detects an incoming signal and exits the scanning loop, thus locking onto the signal of interest. This approach can be seen in Fig. 12.4.

Using controllers, events and parameters in this way, we can build "smart" self-configuring radios which adapt to changes in their operating environment to maintain and optimize communications links.

So far, we have seen how a radio can be created using an XML configuration file and we've seen that a controller can be used to dynamically reconfigure our radio at runtime. Next, we'll look at how to write our own components and controllers to use in our radios.

Listing 12.2 outlines the interface for a Phy component called *ExampleComponent*. To create our own Phy component, we simply create a class which inherits from the PhyComponent class. Within our class, we need to implement a number of functions. Our constructor is the first function required and it takes a string as a parameter. This will be the name which we specify for our component in our XML file. Inside the constructor, we pass this string on to our parent class, along with a description, author and version number. We also use the constructor to register any parameters and events which this component will have.

The second function which we must implement is `registerPorts`. In this function, we register each port which our component will have. We give each port a name and we specify which data types will be supported. Phy components can specify multiple input and output ports, each supporting one or more of a specified set of data types.

The next function we need is `calculateOutputTypes`. This function has two parameters – inputTypes and outputTypes, each a map of port names to data types. The inputTypes parameter tells us which data types to expect on each of our input ports. We then add values to outputTypes to tell the system which data types we will provide on each of our output ports.

The `initialize` function is used to do any setup required by our component. This might include allocating memory or building lookup tables, for example. We are guaranteed that when initialize is called, all of our component parameters will have been correctly set using either our XML configuration or the specified default values. This is not the case for our constructor, hence the need for the additional initialize function. If we need to deallocate memory or do any other cleaning up when our component is unloaded, we can specify a deconstructor and do so there.

The most important function which we define for our component is `process`. This is where the work gets done. Typically, in the process function, a Phy component will read data from the buffer of one or more of its input ports, do some processing on it and then write data to the buffer of one or more of its output ports.

Listing 12.2 Example Phy component interface

```
class ExampleComponent
    : public PhyComponent
{
    public:
    ExampleComponent(std::string name);
    void registerPorts();
    void calculateOutputTypes(
        std::map<std::string, int>& inputTypes,
        std::map<std::string, int>& outputTypes);
    void initialize();
    void process();
};
```

Listing 12.2 outlines the interface for a Stack component in the Iris framework. This time, our component inherits from the StackComponent class. As in a Phy component, the Stack component constructor takes a string parameter which will be passed to the parent class along with a description, author and version. In the constructor we can register the parameters and events of our component.

The `initialize` function is where we do any setup needed by our component and if we require them, we can also define `start` and `stop` functions. These are called when our component is started and stopped respectively and can be used for example, to manage additional threads which our component will run. Stack components define two functions for processing data, according to the source of that data. Data coming from above is processed in `processMessageFromAbove` and data from below is processed in `processMessageFromAbove`. Each function receives the data to be processing as a *StackDataSet* parameter. The StackDataSet contains a set of data, along with metadata such as a timestamp. All data processed by Stack components is of the *uint8_t* type, an unsigned 8-bit integer, so we do not need to specify data types for our component ports.

Listing 12.3 Example Stack component interface

```
class ExampleComponent
    : public StackComponent
{
public:

    ExampleComponent(std::string name);
    void registerPorts();
    void initialize();
    void start();
    void stop();
    void processMessageFromAbove(
        boost::shared_ptr<StackDataSet> set);
    void processMessageFromBelow(
        boost::shared_ptr<StackDataSet> set);
};
```

We've seen how Phy and Stack components can be implemented for the Iris framework. The last module type which we might want to implement is a controller. Listing 12.4 outlines the interface for an example controller. As with components, controllers are implemented as classes which inherit from a Controller base-class. We implement a constructor and pass a name, description, author and version number to the parent class. The `subscribeToEvents` function allows our controller to subscribe to specific events on specific components in our radio. By subscribing, we ensure that our controller will be notified whenever that event is activated. The `processEvent` function is called whenever such an event is activated. In this function we can take whatever action is needed, for example reconfiguring the running radio. We also define `initialize` and `destroy`

functions to do any initial setup or final cleanup which might be needed. Controllers run their own threads of execution and these functions are guaranteed to be called by our controller's own thread.

Listing 12.4 Example controller interface

```
class ExampleController
    : public Controller
{
public:

   ExampleController();
   void subscribeToEvents();
   void initialize();

   void processEvent(Event &e);
   void destroy();
};
```

Iris is available under the open-source GNU Lesser General Public License (LGPL) version 3. See http://www.softwareradiosystems.com/redmine/projects/iris for more information and to download the code [31].

12.4.1.2 GNU Radio

GNU Radio [28] is a free and open-source signal processing toolkit for the implementation of software-defined communication systems. GNURadio was firstly developed and made available for the GNU/Linux operating system, but today it has become a truly cross-platform project, as long as it can be executed on most major operating systems. Even some less common processor architectures like those based on the ARM processor or the PlayStation equipped with Cell processors have been successfully used to run GNU Radio. The following discussion refers to a GNU Radio environment installed on a GNU/Linux system.

The GNU Radio framework is aimed at the definition and handling of complex signal processing structures through the use of predefined blocks available in the GNU Radio library or of custom blocks developed by the user. When a front-end RF device such as the Universal Software Radio Peripheral (USRP) is connected to the workstation, a full transmission system can be realized with the signal-processing part entirely implemented in software. GNU Radio can be used to perform simulations as well, although this is not its main aim. In this case, no RF front-end is needed. Indeed, being aimed at real implementations, it lacks some features for performance assessment that would be useful for simulations.

The GNU Radio environment is based on a layer of signal processing blocks which belong to the *gnuradio-core* library. Blocks are defined by means of C++

or Python [59] classes. Other classes are used to specify flow graphs, the run-time components that schedule block execution, signal buffers, and other basic elements of the GNU Radio system. The Simple Wrapper and Interface Generator (SWIG) [60] library is used to make the C++ classes accessible from a higher-level environment written in Python. In this environment, modules are connected to form a flow graph that defines in software the whole transceiver. Flow graph instantiation and execution is performed by such high-level Python process or by an executable developed for such purpose in C++. A graphical application called GNU Radio Companion (GRC) provides an intuitive user interface for the development of systems using flow graphs. GRC uses the eXtensible Markup Language (XML) to describe both the flow graph and the interface of each of the signal processing modules. The low-level signal processing functions of each block are developed in C++ for maximum efficiency. GNU Radio can be installed following the detailed instructions given on the web site. A first step consists in specifying a simple system using GNU Radio Companion and using it to execute a simple simulation. With GRC, flow graphs can be defined and the corresponding Python code can be automatically generated. For example, typing in a terminal shell the following command

```
$ gnuradio-companion <install path>/share/gnuradio/examples/grc/simple/
variable_config.grc
```

opens a GRC interface with a small example of a flow graph containing a virtual signal generator. The user can modify the parameters of the generated wave and observe the effects of such changes. A XML file (shown in Listing 12.1) is used to specify the block parameters and its input-output ports. Such file is parsed by GRC; the extracted information is used to generate the graphical representation of the block instance in the GUI and to connect such block to others in the flow graph by generating the corresponding Python code.

Once the use and connection of blocks is well understood, one can start to create her own user-defined blocks. For the creation of a new standard block conforming to the GNU Radio conventions, the utility *gr_modtool* can be used. *gr_modtool* is a script that generates a source tree containing the basic structure of a GNU Radio block, including the files needed to compile it and install it (*Makefile*'s and the configuration files), which will be necessarily modified according to the requirements of the project at hand [29]. *gr_modtool* has been included in the GNURadio source tree starting from version 3.6.4, and it is installed by default.

We describe in further detail the procedure that generates a new signal processing block, its Python and XML interface description through *gr_modtool*: from theak

directory where the source tree of the new block should be created, the following command

```
$ gr_modtool newmod mymodule
```

generates a new *module* (in GNU Radio terminology, a module is a set of blocks) structure under subdirectory `gr-mymodule`. It contains eight subdirectories (*apps, docs, grc, cmake, include, lib, python, swig*) and the *CMakeLists.txt* file. The *include* folder contains the C++ public header file of each block, while in the *lib* folder, the C++ private implementation header file and source code corresponding to each block are found. In the *swig* folder, we find the wrapper code that permits to call the public C++ functions and methods (included in the public header file) from Python. The *GRC* folder contains the XML files that describe the interfaces and parameters of each block. A block can be added to the current module using the ***gr_modtool add*** command, a procedure that requires answering a number of questions on the command line about its characteristics. Several block types are available: *sink*, *source*, *general*, *interpolator*, *decimator* and *hierarchical*. When a new block is created, one of these types must be indicated. A source block has one or more outputs and no inputs. Vice versa, a sink block consumes all the data coming through it input port(s), while generating no output data. A general block has both input and output ports. All valid flow graphs contain at least one source block and at least one sink block.

Multiple blocks can be added to each module using ***gr_modtool add*** repeatedly, although in our simple example we will generate a module with a single generic block. After creating the generic block, input and output ports have to be specified according to the design requirements. In the C++ files, the *io_signature* identifiers define the input/output port data types and sizes: these have to be set to proper values. A code segment containing such *io_signature* identifiers follows:

Listing 12.5 Example of how to define inputs and outputs of a GNU Radio block

```
Example_block_impl::Example_block_impl()           :           gr_block("Example_block",
gr_make_io_signature(<+MIN_IN+>,<+MAX_IN+>, sizeof(<+float+>)),
gr_make_io_signature(<+MIN_IN+>,<+MAX_IN+>, sizeof(<+float+>)))
{}
```

The first call to `gr_make_io_signature` specifies the input signature, where the $< +MIN\_IN+ >$ e $< +MAX\_IN+ >$ strings have to be replaced by values indicating, respectively, the minimum and maximum number of input ports. The second call to `gr_make_io_signature` specifies the output signature

with the same syntax. The XML file in the *GRC* subdirectory must be modified coherently with the specification given in the calls to `gr_make_io_signature`.

Listing 12.6 Example XML GRC block configuration

```
<block>
    <name>Example_block</name>
    <key>newblock__Example_block</key>
    <category>newblock</category>
    <import>import newblock</import>
    <make>newblock.Example_block()</make>
    <!-- Make one 'param' node for every Parameter Sub-nodes:
        * name
        * key (makes the value accessible as $keyname)
        * type -->
    <param>
      <name>...</name>
      <key>...</key>
      <type>...</type>
    </param>
      <!-- Make one 'sink' node per input. Sub-nodes:
        *name (an identifier for the GUI)
        *type
        *vlen
        *optional (set to 1 for optional inputs) -->
    <sink>
      <name>in</name>
      <type><!-- e.g. int, complex, vector..--></type>
    </sink>
      <!-- Make one 'source' node per output. Sub-nodes:
        *name (an identifier for the GUI)
        *type
        *vlen
        *optional (set to 1 for optional inputs) -->
    <source>
      <name>out</name>
      <type><!-- e.g. int, complex, vector..--></type>
    </source>
</block>
```

After the block interfaces have been specified both in the C++ and in the XML code, we are ready to write the signal processing code. This code can be split in three parts:

- Resource allocation and initialization;
- Runtime signal processing;
- Resource deallocation.

Resource allocation and initialization code belongs to the constructor of the newly created class. It typically contains memory allocation ad initialization of data structures with preliminarily computed data. Other resources may be allocated as well, such as file handlers or other OS facilities. This code is executed once at the time of instantiation of each block.
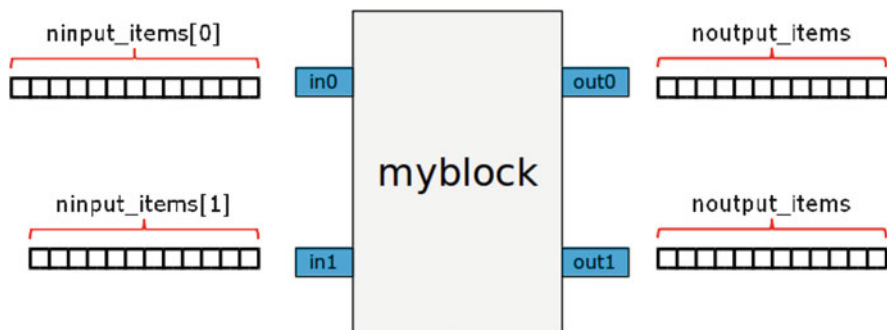
Fig. 12.5 Example of a block with two inputs and two outputs

Resource deallocation code belongs to the destructor of the newly created class. It typically contains deallocation of dynamically allocated memory and of other resources like, e.g., file handlers or other OS facilities. This code is executed once at the time of block destruction.

The *runtime signal processing* code, instead, is executed multiple times throughout the life of each block and determines its runtime behaviour and performance. It belongs to the `general_work()` function of our *Example_block* class. Two arguments of the `general_work()` functions are *noutput_items* and *ninput_items*. As shown in Fig. 12.5, *ninput_items* is a vector and indicates the total number of items available in each input buffer, while `noutput_items` indicates the total number of output items that can be written to each output buffer. `noutput_items` is an integer and not a vector, since in GNU Radio all output buffers have the same size. *input_items* (*output_items*) is a vector of input (output) buffers, where each element corresponds to an input (output) port.

Listing 12.7 General work example

```
int Example_block_impl::general_work (
        int noutput_items,
        gr_vector_int &ninput_items,
        gr_vector_const_void_star &input_items,
        gr_vector_void_star &output_items)
{
    const float *in = (const float *) input_items[0];
    float *out = (float *) output_items[0];
    ...
}
```

In the simpler case of synchronous blocks, a simplified version of runtime processing function called `work()` is defined in place of the `general_work()`. It has a reduced set of input arguments: the `ninput_items` vector is not provided,
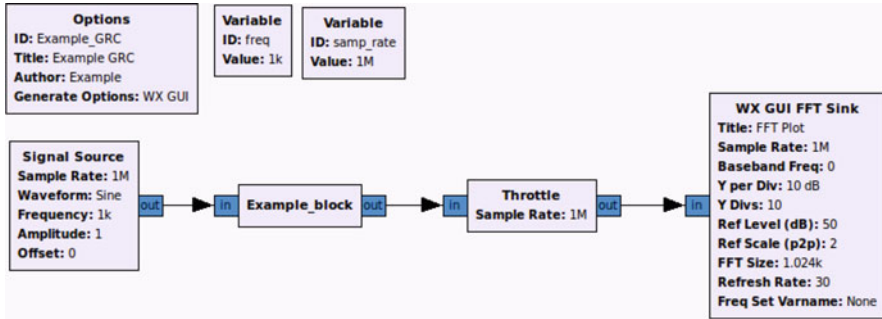
Fig. 12.6 Example of flow graph shown in GRC

since it is assumed that, for synchronous blocks, the same buffer sizes are used both for inputs and outputs.

After compiling and installing the module, the new block can be connected to other blocks to form a flow graph using C++, Python, or gnuradio-companion. The latter is the easiest way, because a flow graph can be created in few steps using a simple and intuitive graphical user interface. This tool is also able to generate the corresponding Python code.

In python you can connect the source and the destination with the *connect* method.

Listing 12.8 Example Python block connection

```
self.connect (src0, (dst, 0)) self.connect (src1, (dst, 1))
```

Listing 12.9 shows the Python code generated by the GRC file shown in Fig. 12.6.

Listing 12.9 Example of connection blocks generated with gnuradio-companion

```
self.Add(self.wxgui_fftsink2_0.win)
self.newblock_Example_block = newblock.Example_block()
self.gr_throttle = gr.throttle(gr.sizeof_gr_complex*1, samp_rate)
self.gr_sig_source = gr.sig_source_c(samp_rate, gr.GR_SIN_WAVE, freq, 1, 0)
################################################
# Connections
################################################
self.connect((self.gr_sig_source_x, 0), (self.newblock_Example_block, 0))
self.connect((self.newblock_Example_block, 0), (self.gr_throttle, 0))
self.connect((self.gr_throttle, 0), (self.wxgui_fftsink2, 0))
```

After the Python code has been generated, it is possible to execute it within gnuradio-companion. This is the point where a hidden entity, the GNURadio scheduler, comes into play. This entity is responsible for the allocation of processing resources across the blocks within a flow graph. There are two types of scheduler: single-threaded scheduler (STS) and thread-per-block (TPB) scheduler; each of

them can be selected by setting the environment variable *GR_SCHEDULER* respectively to 'STS' or 'TPB'. When the single-threaded scheduler is selected, it allocates only one thread for each flow graph so that the runtime signal processing functions are executed sequentially rather than in parallel. The TPB scheduler, instead, assigns to each block a different thread, resulting in parallel execution of blocks. The efficiency of execution results greatly increased, particularly on multi-core processors.

Concerning the scheduler, a GNURadio flow graph can be in one out of the four following different states: *start*, *run*, *stop* and *wait*. When the flow graph is started through the `run()` or the `start()` functions (the former has a blocking behaviour, while the latter does not), the scheduler calls each block executor and polls periodically all of them to verify which one has enough input items and an available output buffer. When this occurs, the corresponding block starts its runtime processing. Iterated polling follows in the same order from source to sink until a `wait()` or a `stop()` method is called. The `wait()` call puts the flow-graph in a stand-by mode and lets it be reconfigured so that blocks and/or connections between them can be changed. After a non-blocking `start()`, the `stop()` method forces the end of execution. In the TPB scheduler mode, each block thread loops continuously over the `work()` method that defines the runtime signal processing function.

12.4.1.3 ASGARD

ASGARD is a framework for building SDR using a set of libraries written in C++ that runs in the Linux environment. The ASGARD API relies on the Poco [45], Boost [7], and Intel Threading Building Blocks [33] libraries, which are used to define the basic functionality of the framework. The framework consists of four basic building blocks; Components, Modules, Communication Objects(CO) and Applications.

Communication Objects are the data containers of the ASGARD framework and implements methods to manipulate the internal data, depending on which CO is used. By implementing thread safe operations and events in the CO data handling methods, it is possible to use them to control the system dataflow in a multi-threaded environment. It is the Component which accesses the data manipulation methods of the CO and this is done through the usage of functors [39], which will be described more thoroughly in the following paragraph. As illustrated in Fig. 12.7, if the implementation of the CO methods supports it, not only is it possible to have multiple Components connected to the same CO but also to the same method. The COs can operate on any desired data type or struct making them very versatile.

The *Component* is the container for implementation of a data processing task and its main purpose is to define the interface around this by handling; IOs, initialization, reconfiguration, and executing. ASGARD follows closely the principle behind
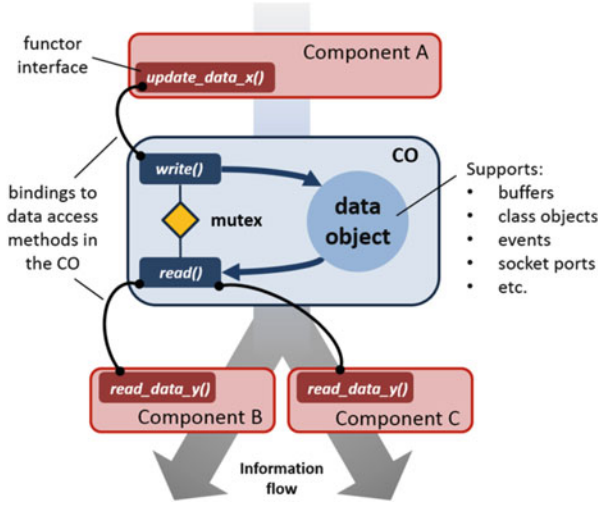
Fig. 12.7 Example of the connection of multiple components through a communication object (CO)

object oriented programming by making each Component a separate class which inherits from the ASGARD Component API. The processing task is specified in the `Do()` method of the Component which is executed by the Module in runtime. The philosophy is that each processing task is executed once for every time the `Do()` method is run. The Component has three important includes which is required; the ASGARD API include Component, the *boost::function* for the functor support, and the *poco::abstractconfiguration* for reading the settings file.

The I/O Ports of a Component is usually specified in the header file:

Listing 12.10 Defining the I/O of an ASGARD component

```
boost::function< void(int&) > read_data_x;
boost::function< void(string&) > write_a_string;
```

The functions defined above can now be used as: `read_data_x(<int>&)` and `write_a_string(<string>&)` in the `Do()` method of the Component.

To enable the possibility of altering settings of a Component without the need for recompiling, the designer can read settings from an XML file in the initialization. This reading is done in the constructor of the Component. As seen later in the description of the Application development, the designer gives each Component a name which it will refer to when checking the settings XML file. The Component name is also used in added standard terminal print messages which can either be

errors, normal print, warnings and more. Below is an example on how ASGARD
abstracts the user from registering these settings.

Listing 12.11   The ASGARD initialization procedure in the constructor

```
//Register the component name to reference in XML
InitializeSettings(string component_name);

// Register settings internally
RegisterSetting(<string> setting1_name,
    <any type> data_variable1,
    <string> default_setting1,
    <ADT> data_type1);
RegisterSetting(<string> setting2_name,
    <any type> data_variable2,
    <string> default_setting2,
    <ADT> data_type2);

// Read the data from the file
LoadSettings(poco::AbstractConfiguration configuration)
```

Multiple setting data types are supported by ASGARD, from standard C++
types, arrays created of vectors to big data structs. In order to easily cope with
this, the settings are always written/read as a string which is also illustrated in the
default_setting variable. However, this sets the need for the designer to remember
which datatype is used. If the data is not found, or if anything goes wrong, this
interface will plot a detailed error such that the designer knows what is wrong
and can easily fix it. There is also a settings debug print which shows all the read
settings.

The Do(), method of a Component is usually structured as seen below.

Listing 12.12   The content of the ASGARD Do() method

```
Do()
{
    // Check for new settings if need for runtime reconfigurability)
    TryReconfigure();
    // Check if component is initialized
    if (!initialized_)
        // initialize settings as desired

    // Get data
    int data_in;
    read_data_x(data_in);

    //manipulate data and print a bit
    ComponentPrint("Manipulating data");

    // set output
    write_a_string() }
```
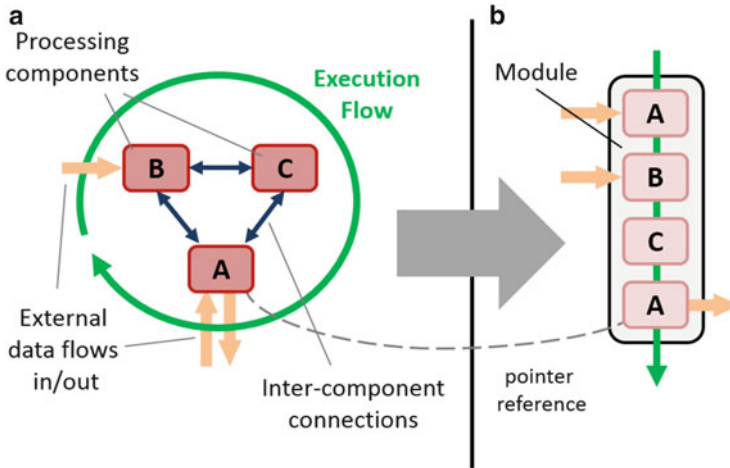
Fig. 12.8 Mapping of a system execution flow (**a**) into a module object (**b**)

As seen above, the `Do()` method uses the functors previously defined to get and set data. The Component does not care about the actual operation of the connected method that it should accept the correct input type. This enables the flexibility of connecting Components to different COs, depending on the desired functionality.

The ***Module*** controls the runtime operation. It is an abstraction which helps the designer managing multiple threads and parallel execution flows in a system application. A Module is fundamentally a list of memory pointers to loaded Components `Do()` method, encapsulated by a threaded object. To load a Component onto a Module, the `AddComponent` method is called as will be shown the Application development section below. When the Module is started it creates a thread which continuously runs the Components `Do()` methods in the order they are loaded. This means that it is also possible to load the same module twice in a row if such operation is desired, as shown in Fig. 12.8.

The ***Application*** is a high-level container which holds the executional part of the code. In the ASGARD Application space, Components, Modules and COs are configured, interconnected and executed.

The ASGARD Application gives the designer full control of the data flow by using suited CO's. The ASGARD frameworks flexibility comes from the ability to choose between a variety of data connections an container methods. This flexibility enables support of high level CR/DSA solutions using cross layer Component communication. However, the flexibility comes at the risk of low stability on poorly designed systems due to unexpected behaviour. In order to minimize the risk of error due to bad design, the ASGARD software encourages the use of Test Driven Development [5] throughout the entire development process.

Listing 12.13 A simple application example of ASGARD

```cpp
class ExampleApp
{
   ExampleApp()
   {
      // Open and load the configuration file to memory
      Configuration config = loadConfiguration("settings.xml");

      // Instantiate Components
      component1_ = new ComponentClass1("Component1", config);

      // Load Components to Modules;
      module1_.AddComponent(component1_);

      // Connect the Component methods to the CO
      component1_->read_data = boost::bind(&Buffer<int>::pop,
         boost::ref( co1_ ), _1);
      component1_->write_string = boost::bind(&Buffer<string>::push,
         boost::ref( co2_ ), _1);

      // Set settings available to the other Components
      component1_->Initialize()
   }

   main(const std::vector< std::string >& /* args */)
   {
      // Start module(s)
      module1_.Start();

      // Wait for user input to stop
      module1_.Stop();
   }

   // Define Component pointers, COs and Modules
   Module module1_;
   ComponentClass1* component1_;
   Buffer<int> co1_;
   Buffer<string> co2_;
};
```

ASGARDs design philosophy also enables easy implementation of cross-platform building blocks, extending the optional Component library to span over a wide variety of Components. This enables the usage of features from multiple frameworks in the same application, but using ASGARDs versatile philosophy.

12.4.2 Running Experiments: Best Practices

This section will contain a description of experiments which has been executed using the frameworks described in Sect. 12.4.1. This is to give the user an idea on how the theory described through the chapter can be used to design a test bed and to give insight in the process.

12.4.2.1 Multi Carrier PHY Enhancement with IRIS

In wireless telecommunications research, simulations can play an important role in assessing the value of research outputs, from novel waveform designs, signal detectors and synchronizers to MAC protocols and rendezvous mechanisms. In order to truly prove the value of research output however, experimental analysis using live over-the-air waveforms is often needed. In this section, we provide an overview of a set of experiments carried out on the Trinity College Dublin software radio test bed using the Iris architecture.

The set of experiments we will focus on were carried out to examine the performance of a novel PHY layer technique for achieving rendezvous in dynamic spectrum access systems [49–51]. The technique involves the use of *cyclostationary signatures*, physical watermarks that can be embedded in multicarrier waveforms and used to assist signal detection, identification and carrier frequency estimation. The advantage of the technique is that it allows these processes to be carried out at the receiver with minimal knowledge of the parameters of the transmitted waveform.

Three experiments were designed. The first examined signal detection performance, the second signal identification and the third examined carrier frequency estimation. In each case, the impact of varying a range of parameters on both the transmitter and receiver were assessed. The experiments closely matched simulations which had been carried out using the Matlab environment and could thus be used to assess and verify those simulation results. In each case, the experiments required a single transmitter and a single receiver. They were carried out indoors, in a test bed environment and were automated to run overnight when activity in the test bed was minimal. Test licenses were obtained from Comreg, the Irish communications regulator for a 25 MHz band centred at 2.35 GHz. This frequency band was used for all experiments.

The equipment used in the experiments included an Anritsu MG3700A vector signal generator and an Anritsu MS2781A Signature spectrum analyzer. The MG3700A allows arbitrary waveforms to be generated and transmitted over the air at frequencies between 250 kHz and 6 GHz and with sample rates of between 20 kHz and 160 MHz. The MS2781A supports frequency ranges between 100 Hz and 8 GHz with resolution bandwidths between 0.1 Hz and 8 MHz. The analyzer was not used to receive signals of interest but was used extensively to verify carrier frequencies, power levels and bandwidths of over-the-air signals. In addition, Ettus Research LLC Universal Software Radio Peripheral (USRP) hardware was used as the RF front-end for our Iris software radio architecture. The USRP2 front-end transfers baseband I/Q samples over a gigabit ethernet interface to and from a PC for baseband processing. In this way, arbitrary waveforms can be transmitted and received within a 25 MHz RF bandwidth (at 16 bit sample resolution) at frequencies between DC and 6 GHz.

In our first experiment, we examined the performance of a signal detector using embedded cyclostationary signatures. Here we used the MG3700A signal generator together with a single USRP front-end and our Iris software radio architecture. OFDM signals were generated with a range of parameters using and Iris transmitter

configuration and stored to file in baseband I/Q format. These were then loaded onto the signal generator and transmitted at 2.35 GHz with a 1 MHz bandwidth. Transmitted signals were captured at a distance of 3 m using the USRP front-end with an RFX2400 daughterboard and transferred over the GigE interface to the Iris software radio and processed by a receiver radio configuration. Transmit power levels were set on the MG3700A to between −8 and −36 dBm to give an estimated receive SNR of between 12 and −6 dB. SNR estimation at the receiver was achieved using captured signals comprising noise only and both noise and a known OFDM signal. The PSD was calculated for each and averaged over 1,000 windows of 1,024 samples. Mean powers were estimated over bandwidths of both the noise only (P_n) and the noise and signal samples (P_{s+n}). These mean powers were then used to calculate the estimated SNR as:

$$SNR_{est} = 10 log_{10}(\frac{P_{s+n}}{P_n} - 1) \tag{12.1}$$

For each parameter value, 1,000 experiment runs were carried out and results were captured to file for post-processing. In order to assess the detector performance, signals both with and without embedded signatures were used. From recorded results, it was possible to perform ROC analyses which examine the achievable probability of detection and probability of false alarm for the full range of usable detector thresholds.

For our second experiment, we examined the performance of cyclostationary signatures when used for signal identification. Here, the experimental setup was similar to that of our first experiment. Suitable signals were generated and loaded onto the MG3700A signal generator. These were then transmitted at 2.35 GHz and received using the USRP front-end and an Iris radio. A range of parameters were chosen for the transmitted waveforms and 1,000 experiment runs were used for each. Results were recorded to file for post-processing. Using this approach, it was possible to examine the signal identification performance over estimated SNR values between −5 and 10 dB for a wide range of signature types. Figure 12.9 illustrates the Iris radio configuration used in our experiments. A single PHY engine was used with a USRP receiver component and a cyclostationary analyzer component. The analyzer carried out cyclostationary signature detection at a single cyclic frequency and provided detection results to an Iris controller. Different controllers were used for each experiment, according to the specific measurements which needed to be captured. In all cases, results were also recorded to file for later analysis.

Our third experiment examined the performance of cyclostationary signatures when used for carrier frequency estimation. Again, the experimental setup included the MG3700A signal generator, a USRP front-end and the Iris software radio architecture. However, for this experiment it was necessary to randomly vary the carrier frequency of the transmitted signal within a predefined frequency band. Programming scripts were used to automate the testing process.
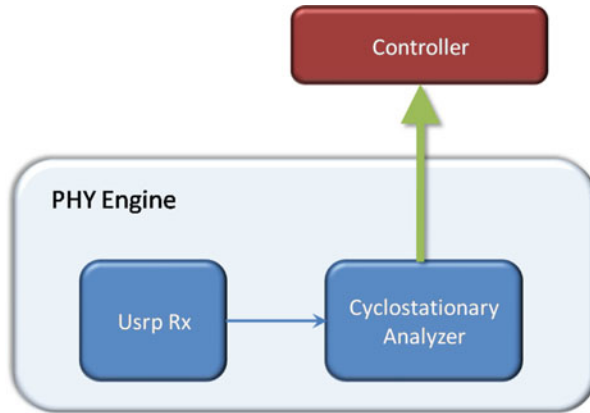
Fig. 12.9 Iris configuration used for experiments
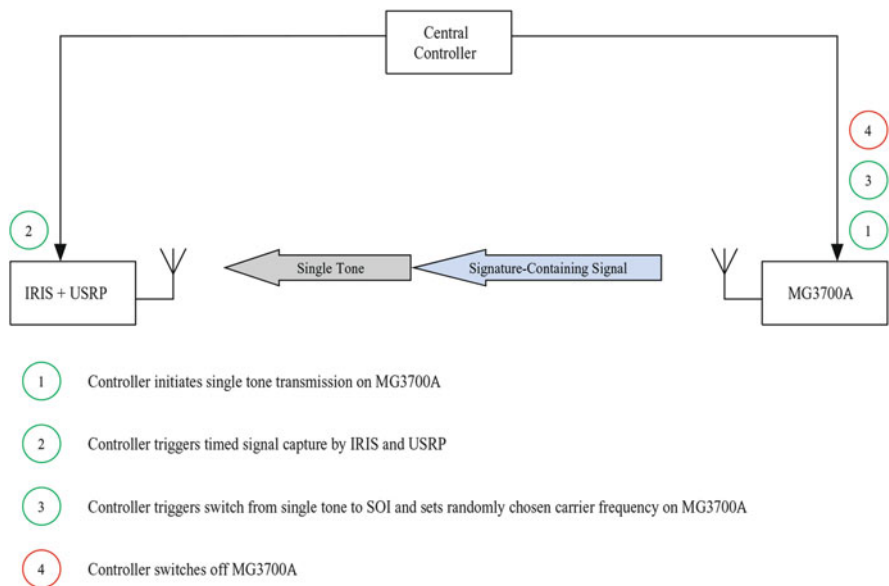
Fig. 12.10 Automated signal capturing sequence used for frequency acquisition tests

Figure 12.10 illustrates the sequence used for each test. Remote control of the signal generator was facilitated using an ethernet interface while the Iris software radio engine was dynamically reconfigured using XML configuration scripts, pushed to Iris over ethernet. The sequence for each test comprised four steps. Firstly, the central controller initiated transmission of a single tone at 2.35 GHz. This tone was used to calibrate the USRP front-end for each test. Secondly, the controller initiated a timed signal capture on the Iris radio. The first part of this capture consists

of the single fixed tone. Once the tone was captured, the controller triggered the MG3700A to switch the transmitted waveform from the single tone to the OFDM signal containing the embedded signature. At the same time, the carrier frequency of that signal was adjusted to a randomly chosen frequency between 2.2485 and 2.3515 GHz. Once that signal had been captured using Iris, the controller switched off transmission on the signal generator to complete the test. In this way, each signal capture comprised two parts, the first containing the single tone and the second containing the SOI at a carrier frequency randomly chosen by the controller.

This signal capture sequence was performed repeatedly, in a loop, to provide sufficient captured data to perform our analysis. Following the signal capture process, the signal data for each test was processed offline using our transceiver implementation to perform carrier frequency estimation and provide results. A probability of signal acquisition was defined and examined for a range of estimated SNR values.

For each of our experiments, we used the available laboratory equipment to maximize accuracy and minimize the possibility of error. In many cases, it is necessary to calibrate RF hardware using laboratory instruments and, as seen in our third experiment, it may be possible to do this online, during the experiment itself. For over-the-air experiments, it is often useful to first carry out tests over coaxial cable between the transmitter and receiver to limit the unexpected effects of other nearby RF emitters. Once the tests have been verified over a coaxial link, they can then be carried out over the air.

12.4.2.2 TV White Spaces Detection with GNU Radio

There is a great interest around the development of efficient spectrum sensing algorithms to be used in the TV White Spaces (TVWS) for the identification of primary user signals and possible exploitation of available channels. The TVWS are channels in the part of UHF band used for TV broadcasting (470–862 MHz) in which primary users (i.e., TV broadcasters) do not exploit their right of use by not performing transmissions in certain areas. Secondary users could dynamically allocate spectrum for broadband communications in those channels. Primary users are TV broadcasters that transmit signals conforming to the DVB-T standard [15] in which coded OFDM is used.

Spectrum sensing of coded OFDM signals can be performed using several techniques: the DVB-T signal can be efficiently detected using the so-called "feature-based" detectors, which aim to exploit signal parameter knowledge in order to achieve a rather high sensitivity.

We developed a test-bed wherein a dedicated spectrum sensing unit is able to sense the DVB-T channel through the USRP2, process the received signal and verify the presence of the Cyclic Prefix (CP). The CP represents an example of *signal feature* embedded into almost all OFDM signals. The DVB-T signal exhibits many other potential signal features such as, for example, scattered and continual pilot symbols and repeated Transmission Parameter Signalling (TPS) code words.

A good trade-off between detection efficiency and complexity is achieved using the *lagged autocorrelation*, a mathematical tool of wide use in signal analysis. This tool relies on the cyclostationary properties of the signal to be detected in order to compute a test statistics that, compared with a threshold value, implements the binary hypothesis testing setting, common to all spectrum sensing devices.

In [6], the detectors based on CP autocorrelation described in [12] have been implemented, improved, and applied to a real scenario using GNU Radio. The implemented algorithm computes the accumulated autocorrelation function over K OFDM symbols:

$$R_{xx}^{(CP)}[\tau] = \frac{1}{KN_s} \left| \sum_{k=0}^{K-1} \sum_{n=\tau-kN_s}^{\tau-kN_s-N_{CP}+1} x^*[n]x[n-N_{FFT}] \right|. \qquad (12.2)$$

In (12.2), K represents the total number of OFDM symbols over which the autocorrelation is accumulated before taking the absolute value. Each OFDM symbol consists of $N_s = N_{CP} + N_{FFT}$ samples, where N_{CP} is the number of CP samples and N_{FFT} is the number of data samples.

In order to perform CP-based spectrum sensing, it is necessary to define a threshold to compare with the computed autocorrelation values. If $R_{xx}^{(CP)}[\tau]$, for some τ, exceeds such threshold, a primary signal is assumed to be present. A typical approach consists in setting this threshold as a function of the average noise power, if known, and the desired false alarm probability (the probability that the detector indicates the presence of a primary signal when it is not present). In this case, a noise estimation algorithm is needed. Without any a-priori assumption, the noise estimation should be performed by analysing the received samples in absence of primary signal. However, this approach is often impractical and prone to noise variations.

The approach proposed in [6] considers the peaky shape of the autocorrelation and computes the threshold by averaging its samples in a well-defined window between two consecutive correlation maxima. As output it is possible to obtain the autocorrelation function and the computed threshold of the GNURadio block used: the OFDMsense block. As input to the OFDMsense block is a sample stream from a USRP device (interlaced IQ values) on which the block calculates the test statistics.

The block parameters are N_{FFT} (number of data samples), N_{CP} (number of CP samples) and a scaling factor used to compute the threshold value starting from the sample average of autocorrelation.

The performance of the proposed feature-based spectrum sensing algorithm have been evaluated considering different numbers of consecutive OFDM symbols ($K = 1, 10, 100$ symbols). The primary signal consist of a real DVB-T signal: 8k carriers, CP length ratio of 0.25, channel bandwidth of 8 MHz (a typical configuration). Excellent results in terms of detection probability at very low signal-to-noise ratio (SNR) values have been obtained, as shown in [6].

DVB-T channels can be processed using the USRP2 or the newer USRP N200 series of SDR peripherals. The USRP2 is capable of sampling a received signal at

100 MHz (quadrature sampling), with a precision of 14 bits/sample and stream the IQ samples to the host after decimation with a sample rate of 25 MHz through its 1 Gbps Ethernet interface. In our test bed, the USRP2 was set to process a bandwidth of 12.5 MHz (that corresponds to an FPGA decimation factor of 8) with a single 8 MHz DVB-T channel.

12.4.2.3 Dynamic Spectrum Access with ASGARD

Future generation wireless systems will assume access points (APs) will be deployed by the users himself in indoor office and residential environments. In such context, where planned reuse of frequencies between the APs is infeasible, Dynamic Spectrum Access (DSA) systems represent a valid alternative for optimizing spectrum resource utilization and ensuring minimization of interference. At Aalborg University a series of experimental trials have been carried out with the Autonomous Component Carrier Selection (ACCS) [24], a DSA algorithm for interference mitigation purposes. On the contrary of many link-level focused solutions, the ACCS is a distributed decision making process with explicit coordination among the network peers. APs in the network are able to exchange high-level information about their occupied spectrum resources and estimations of the interference they generate towards of other cells. Such information exchange occurs over a dedicated control channel. ACCS has been extensively analyzed by prior simulation-based work, the purpose of the experiments was to validate in realistic deployment conditions the prior findings in terms of baseline performance of the ACCS-enabled network.

An SDR test bed was realized for ACCS based on USRP hardware and the ASGARD software platform. The main objective for the test bed design was to develop an easy-deployable and scalable solution which would enable both the runtime execution of a reconfigurable communication system architecture and the management of network-wide operations. Figure 12.11 shows how the desired software architecture of the node was implemented in the ACCS test bed. The experimentation with ACCS focused on the analysis of the network performance in terms of Signal to Interference and Noise Ratio (SINR) experienced by the users in pre-defined network cells. Test bed nodes were deployed in buildings exploiting rooms to physically identify different cells. In the experiments here considered multiple nodes have been placed across a single floor of office premises at Aalborg University. Every cell is constituted by a single AP and one UE, the UE is always placed in the same room of the affiliated AP. Figure 12.12 provides an overview of the ACCS test bed deployment considering 12 nodes and 6 cells.

The ACCS algorithm manages the allocation of transmission resources in each AP assuming the total available bandwidth to be divided in a number of spectrum chunks. The decision making process relies on channel sensing information related to the UEs and/or the APs estimating the interference generated by neighbouring APs. Sensing data is then combined with information from the control channel before ACCS takes a decision about which spectrum resources to allocate. The execution of ACCS is periodic: we assume all APs to have common frame-level
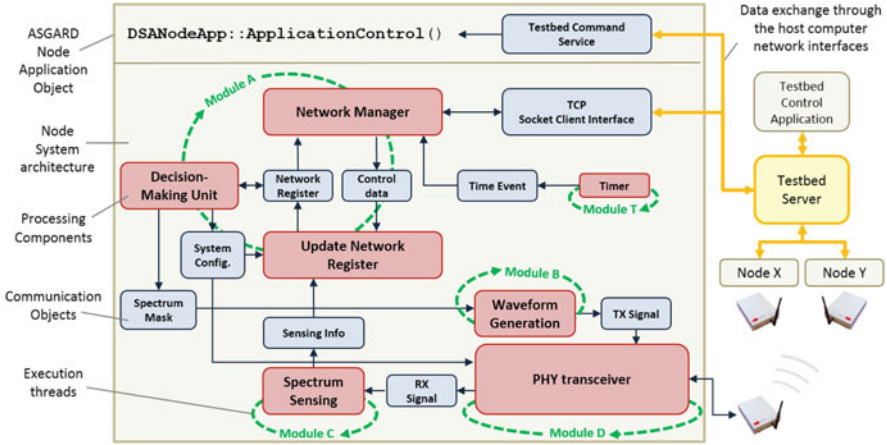
Fig. 12.11 The architecture of the ACCS test bed node

Fig. 12.12 Deployment of nodes of the ACCS test bed across office premises at Aalborg University

synchronization where the duration of a frame is in the order of hundreds of milliseconds.

From a physical layer perspective, the most important function in the ACCS test bed is the measurement of the received signal power in respect to the transmitting nodes. The key issue in this setup is to clearly identify the source of the transmission in order to estimate the SINR and the potential Carrier to Interference (C/I) ratios occurring between nodes allocating overlapping resources. The sensing of all APs

moreover, must be performed within the time of one ACCS frame which in this case is set to 500 ms. A simple solution to this problem has been obtained by employing orthogonal pilot patterns in the APs OFDM-like transmitted signal. Sensing nodes can concurrently measure the power of all patterns by performing a single FFT operation across the entire transmission bandwidth. The operating bandwidth is here set to 10 MHz and it is obtained by exploiting 800 bins of a 1024-IFFT with USRP sample rate at 12.5 MS/s. Resources at the edge of the band are discarded due to the RF imperfections. In the experiments the total reference power is set to 0 dBm and it is achieved by using a analogue transmission gain of 20 dB and receive gain of 30 on the USRPs.

In order to ensure a consistent measurement of the signal power from multiple links, all RF front-ends have been calibrated both in transmission and reception. The calibration procedure is divided in two phases: first a reference board is manually calibrated with a signal generator and a spectrum analyzer, subsequently all other boards are calibrated from the reference by means of an automated software application specifically developed for this purpose. The developed ACCS test bed employs USRP N200/N210 motherboards together with the XCVR 2450 RF daughterboards. The SDR hardware is connected to host computers featuring Intel i7 core processors, 8 GB RAM and solid state HDs. In order to avoid sources of external interference, all experiments have been carried out in the 5 GHz band which proved to be unused.

The experiments with ACCS consider the bootstrap of the network and the live execution of the DSA process for predefined time periods. SINR measurements from the users in the cells are analysed in order to evaluate the interference mitigation capabilities of the algorithm. Different trials have been conducted in both static propagation environment conditions (running the experiment during night time) and dynamic (thus allowing human presence in the buildings). Multiple experimental runs have been performed: in static environment conditions the experiments have been repeated across a range of carrier frequencies (from 4.91 to 5.89 GHz) in order to provide averaging to the effect of multipath fading. Variable configuration parameters such the number of spectrum chunks available for allocation have also been investigated in independent runs (about 1,400 runs have been conducted during the entire ACCS campaign). In order to manage the execution of a large number of experiments with reconfigurable system parameters and multiple nodes automatized procedures have been developed. The experimental Application running on each test bed machine has been designed such that it automatically reconfigures the system according to the experimental procedure of interest. In order to handle specific events during the experimental runs (e.g. delayed activation of specific nodes) all applications are able to exploit Network Time Protocol (NTP) synchronization provided by the host PCs. The entire data-exchange plane in the test bed (e.g. control data, UE feedback, results acquisition) runs over a parallel backhaul connection exploiting WiFi and/or Ethernet interfaces. A test bed server provides centralized control for the test bed management and logging of data from all the nodes.

12.5 Conclusion

It is a fact that research in CR is in need of moving towards practical experimentation in order to prove on the field the effectiveness of the methods. For this reason there is an increasing trend, for research groups, in jumping into developing their own test beds. While the number of practical experimentations is increasing, an even wider number of researchers, that would be willing to experiment in real life, is restrained by not being familiar with the test beds world. In this chapter, based on direct and practical experience in developing successful test beds and experiments, we have provided a "how to" guide on how to approach the test bed development tackling both the everyday issues, and the scientific ones. The hope is to help more researchers entering the development world in a smoother way, being aware, from the beginning of what they would face, and allowing them to produce scientific results faster.

References

1. 1900.6-2011: IEEE standard for spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems
2. Abdallah, A.S., MacKenzie, A.B., DaSilva, L.A., Thompson, M.S.: On software tools and stack architectures for wireless network experiments. In: On Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), Cancun, pp. 2131–2136, Mar 2011
3. Aryafar, E., Anand, N., Salonidis, T., Knightly, E.W.: Design and experimental evaluation of multi-user beamforming in wireless LANs. In: Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MobiCom), Chicago, Sept 2010
4. ASGARD: A software framework for the development of SDR and cognitive test beds. Online: asgard.lab.es.aau.dk/
5. Beck, K.: Test Driven Development: By Example. Addison-Wesley Professional, Boston (2003)
6. Benco, S., Crespi, F., Ghittino, A., Perotti, A.: Software defined white-space cognitive systems: implementation of the spectrum sensing unit. In: The 2nd Workshop of COST Action IC0902, Castelldefels/Barcelona, Oct 2011
7. Boost C++ Libraries: http://www.boost.org/ (2013)
8. Borkowski, D., Brühl, L., Degen, C., Keusgen, W., Alirezaei, G., Geschewski, F., Oikonomopoulos, C., Rembold, B.: SABA: a testbed for a real-time MIMO system. EURASIP J. Appl. Signal Process. **2006**, 1–15 (2006). Article ID 56061
9. Caban, S., Mehlführer, C., Langwieser, R., Scholtz, A.L., Rupp, M.: Vienna MIMO testbed. EURASIP J. Appl. Signal Process. **2006**, 1–13 (2006). Article ID 54868
10. Cognitive Radio Experimentation World (CREW) Project: Project overview. Available via CREW website. http://www.crew-project.eu/repository
11. Community Resource for Archiving Wireless Data At Dartmouth (CRAWDAD): Data sets. Available via CRAWDAD website.http://crawdad.cs.dartmouth.edu/data.php
12. 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), Istanbul, Sept 2010
13. Doyle, L.E., Sutton, P.D., Nolan, K.E., Lotze, J., Ozgul, B., Rondeau, T.W., Fahmy, S.A., Lahlou, H., DaSilva, L.A.: Experiences from the Iris testbed in dynamic spectrum access and

cognitive radio experimentation. In: 2010 IEEE Symposium on New Frontiers in Dynamic Spectrum, Singapore, pp. 1,8, 6–9 Apr 2010. doi:10.1109/DYSPAN.2010.5457835

14. Ettus Research: Online: www.ettus.com
15. European Telecommunications Standards Institute: Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television. EN 300 744 v1.6.1, Jan 2009
16. Fàbregas, A.G., Guillaud, M., Slock, D., Caire, G., Gosse, K., Rouquette, S., Dias, A.R., Bernardin, P., Miet, X., Conrat, J.-M., Toutain, Y., Peden, A., Li, Z.: A MIMO-OFDM testbed for wireless local area networks. EURASIP J. Appl. Signal Process. **2006**, 1–20 (2006). Article ID 18083
17. Future Internet Research and Experimentation (FIRE): Available via FIRE website. http://www.ict-fire.eu/home.html
18. FP7-CREW Project: Final deliverable D2.1: definition of internal usage scenarios. Available via CREW website. http://www.crew-project.eu/sites/default/files/deliverables/CREW_D2.1_TUD_R_PU_2011-01-31_final_PRC.pdf
19. FP7-CREW Project: Final deliverable D2.2: definition of the federation. Available via CREW website. http://www.crew-project.eu/sites/default/files/deliverables/CREW_D2.2_TCD_R_PU_2011-03-31_final_PRC.pdf
20. FP7-CREW Project: Final deliverable D3.1: basic operational platform. Available via CREW website.http://www.crew-project.eu/sites/default/files/CREW_D3.1_TCF_R_P_2011-09-30_final.pdf
21. FP7-CREW Project: Final deliverable D4.2: methodology for performance evaluation. Available via CREW website.http://www.crew-project.eu/sites/default/files/CREW_D4.2_IBBT_R_PU_2012-09-30_final.pdf
22. FP7-CREW Project: The CREW repository. Available via CREW common portal.http://www.crew-project.eu/repository
23. FP7-FARAMIR Project: Spectrum use measurements in several European countries. Available via FARAMIR website. http://www.ict-faramir.eu/index.php?id=download
24. Garcia, L.G.U., Pedersen, K.I., Mogensen, P.E.: Autonomous component carrier selection: interference management in local area environments for LTE-advanced. IEEE Commun. Mag. **47**(9), 110–116 (2009)
25. Gollakota, S., Perli, S.D., Katabi, D.: Interference alignment and cancellation. In: SIGCOMM'09, Barcelona, 17–21 August 2009
26. Haustein, T., Forck, A., Gäbler, H., Jungnickel, V., Schiffermüller, S.: Real-time signal processing for multiantenna systems: algorithms, optimization, and implementation on an experimental test-bed. EURASIP J. Appl. Signal Process. **2006**, 1–21 (2006). Article ID 27573
27. http://cornet.wireless.vt.edu/trac/wiki/CORNET
28. http://gnuradio.org
29. http://gnuradio.org/redmine/projects/gnuradio/wiki/OutOfTreeModules
30. http://www.openairinterface.org/
31. http://www.softwareradiosystems.com/redmine/projects/iris
32. EEE Standard digital interface for programmable instrumentation, Institute of electrical and electronics engineers, 1987, ISBN 0-471-62222-2, ANSI/IEEE Std 488.1-1987, p. iii
33. Intel Threading Building Blocks: (Intel TBB): http://threadingbuildingblocks.org/. Intel. (2013)
34. Kurkowski, S., Camp, T., Colagrosso, M.: MANET simulation studies: the incredibles. ACM's Mob. Comput. Commun. Rev. **9**(4), 50–61 (2005)
35. Li, L.E., Alimi, R., Shen, D., Viswanathan, H., Yang, Y.R.: A general algorithm for interference alignment and cancellation in wireless networks. In: Proceedings of IEEE INFOCOM, San Diego, Mar 2010
36. Miranda, J.P.: Multi-standard context-aware cognitive radio: sensing and classification mechanisms. Ph.D. dissertation, Apr 2013
37. Mishra, S.M., Cabric, D., Chang, C., Willkomm, D., van Schewick, B., Wolisz, A., Brodersen, R.W.: A real time cognitive radio testbed for physical and link layer experiments. In: 2005

First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), pp. 562,567, 8–11 Nov 2005. doi:10.1109/DYSPAN.2005.1542670

38. Mitola, J., III, Maguire, G.Q., Jr.: Cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
39. Musser, D.R., Derge, G.J., Saini, A.: STL tutorial and reference guide: C++ programming with the standard template library. In: Musser, D.R., Gillmer J.D., Saini, A. (eds.) STL tutorial and reference guide: C++ programming with the standard template library. Addison-Wesley Professional, Boston (2009)
40. Nieto, X., Ventura, L.M., Mollfulleda, A.: Gedomis: a broadband wireless MIMO-OFDM testbed, design and implementation. In: Testbeds and Research Infrastructures for the Development of Networks and Communities, 2006. TRIDENTCOM 2006. 2nd International Conference on, pp. 10–pp. IEEE (2006)
41. Nutaq (formery Lyrtech): Online: http://nutaq.com/en/products/view/+nutaq-sff-sdr
42. Online: http://www.pervices.com/
43. Pawelczak, P., Nolan, K., Doyle, L., Oh, S.W., Cabric, D.: Cognitive radio: ten years of experimentation and development. IEEE Commun. Mag. **49**(3), 90–100 (2011)
44. Pawlikowski, K., Jeong, J., Lee, R.: On credibility of simulation studies of telecommunication networks. IEEE Commun. Mag. **40**(2), 132–139 (2002)
45. Poco C++ Libraries: http://pocoproject.org/ (2013)
46. Ramírez, D., Santamaría, I., Pérez, J., Vía, J., García-Naya, J.A., Fernández-Caramés, T.M., Pérez-Iglesias, H.J., González-López, M., Castedo, L., Torres-Royo, J.M.: A comparative study of STBC transmissions at 2.4 GHz over indoor channels using a 2 × 2 MIMO testbed. Wirel. Commun. Mob. Comput. **8**(9), 1149–1164 (2008)
47. Rice University: Rice university WARP project, available at: http://warp.rice.edu
48. SORA SDR Platform: Online: http://research.microsoft.com/en-us/projects/sora/
49. Sutton, P.D.: Rendezvous and coordination in OFDM-based dynamic spectrum access networks. PhD dissertation, University of Dublin, Trinity College (2008)
50. Sutton, P.D., Doyle, L.E.: Enhanced low-complexity detector design for embedded cyclostationary signatures. Analog Integr. Circuits Signal Process. **73**(2), 495–502 (2012). ISSN:0925-1030, doi:10.1007/s10470-012-9967-8
51. Sutton, P.D., Nolan, K.E., Doyle, L.E.: Cyclostationary signatures in practical cognitive radio applications. IEEE J. Sel. Areas Commun. **26**(1), 13,24 (2008). doi:10.1109/JSAC.2008.080103
52. Sutton, P.D., Lotze, J., Lahlou, H., Fahmy, S.A., Nolan, K.E., Ozgul, B., Rondeau, T.W., Noguera, J., Doyle, L.E.: Iris: an architecture for cognitive radio networking testbeds. IEEE Commun. Mag. **48**(9), 114,122 (2010). doi:10.1109/MCOM.2010.5560595
53. Tavares, F., Tonelli, O., Berardinelli, G., Cattoni, A.F., Mogensen, P.E.: ASGARD: the Aalborg University cognitive radio software platform for DSA experimentation. Presented at the 3rd international workshop of the COST action IC0902: cognitive radio and networking for cooperative coexistence of heterogeneous wireless networks, Ohrid (2012)
54. The CREW Testbed Federation: Online: www.crew-project.eu/crewtestbed
55. The GNU Radio Software Radio Toolkit: Online: gnuradio.org
56. Tonelli, O., Buthler, J.L.: The Asgard Platform. In: A book (ed) Of platforms, 1st edn. Wiley, New York (2013)
57. Vanka, S., Srinivasa, S., Gong, Z., Vizi, P., Stamatiou, K., Haenggi, M.: Superposition coding strategies: design and experimantal evaluation. IEEE Trans. Wirel. Commun. **11**(7), 2628–2639 (2012)
58. www.crc.ca/coral
59. www.python.org
60. www.swig.org

Chapter 13
Low-Cost Testbed Development and Its Applications in Cognitive Radio Prototyping

Tomaž Šolc, Carolina Fortuna, and Mihael Mohorčič

Abstract Having a huge potential to improve the way radio spectrum is being used, the techniques that are used for the research in cognitive radio are maturing and therefore move from being evaluated in a simulation environment to more realistic environments such as dedicated testbeds. In this chapter we describe our experiences with the design, deployment and experimental use of the LOG-a-TEC embedded, outdoor cognitive radio testbed, based on the VESNA sensor node platform. We describe the choice of experimental low-cost reconfigurable radio frontends for LOG-a-TEC and discuss the potential capabilities of custom designs. The core part of this chapter gives practical experiences with designing the embedded testbed infrastructure, covering topology design and performance evaluation of the management network as well as our considerations in the choice of network protocols employed in the LOG-a-TEC testbed. Finally, we provide two use cases where the LOG-a-TEC testbed has been used for performing experiments with cognitive radio, one relevant to the investigation of coexistence of primary and secondary users in TV white spaces and the other addressing power allocation and interference control in the case of shared spectrum.

13.1 Introduction

As a research area in wireless communications that is fundamentally important for more efficient use of radio spectrum, cognitive radio already surpassed the early phase of theoretical only investigations in controlled environments and computer simulations, and has already entered the phase of experimental research and pilot

T. Šolc (✉) • C. Fortuna • M. Mohorčič
Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
e-mail: tomaz.solc@ijs.si; carolina.fortuna@ijs.si; miha.mohorcic@ijs.si

trials in increasingly realistic operating environments. This has been recognised by a number of institutions and projects worldwide, resulting in numerous testbeds, on one hand trying to be sufficiently representative for actual targeting scenarios, and on the other hand striving to be easy for use and to be able to accommodate as diverse algorithms and approaches as possible. As a result, they are inclined towards the use of high-performance general purpose processors and software defined radio (SDR) platforms. Conventional testbeds that are built around this architecture offer a fast way for researchers to implement and test various protocols. However cognitive radio technologies are also interesting for low-powered and low-bandwidth applications, like wireless sensor nodes and body area networks. Such applications often do not have the possibility of including a software defined radio platform in the final product due to power and cost constraints. While most of the basic research can still be done on conventional testbeds, aspects like power, storage and processing limitations on such systems can only be reliably tested on embedded hardware that more closely resembles realistic platforms such applications would run on. Embedded testbeds can also be more easily deployed in configurations where a conventional approach would be difficult. One example of that are outdoor environments, where operation of delicate laboratory equipment can be problematic. Another example are locations that lack existing infrastructure such as Ethernet wiring or constant power supply. The relatively low cost of embedded devices can also be an important benefit in testbeds consisting of a large number of nodes.

A node in a conventional testbed is typically built around a computer running a GNU/Linux-based operating system with a software defined radio front-end and a TCP/IP based management network. A typical setup consists of nodes with a gigabyte of RAM, x86 CPU in the gigahertz clock range and connection to a gigabit Ethernet network. On the other hand, testbeds built up by embedded devices such as the Tmote Sky [1] or VESNA [2] wireless sensor nodes, are powered by microcontrollers with CPU cores like AVR or ARM with clock frequencies in the 10–100 MHz range and RAM sizes from 1 to 100 kB. Typically communication between the nodes involves low-power narrow-band radios with bulk transfer rates in the range from 100 bytes/s to 100 kB/s. Nodes in such a setup either run specialized, small-footprint operating systems like Contiki [3] and TinyOS [4] or even just bare bone firmware without a proper operating system. Network protocols encountered in such testbeds are for example IEEE 802.15.4, ZigBee and Bluetooth. Such systems offer very little built-in management functionality that is usually expected from conventional, networked multi-user computers. Therefore the design and use of embedded testbeds pose several additional challenges over those found in conventional testbeds.

In this chapter we describe our experiences with the design, deployment and experimental use of the LOG-a-TEC embedded, outdoor cognitive radio testbed, based on the VESNA [2] sensor node platform, which is integrated with the CREW federation [5] of cognitive radio testbeds. Where possible, we also mention other approaches to design challenges that were taken in other similar testbeds. We describe the choice of experimental radio frontends for LOG-a-TEC and discuss

the possibility of custom designs. We also give practical instructions for designing the testbed infrastructure, like the management network and our considerations in the choice of network protocols employed in the LOG-a-TEC testbed. Finally, we provide two use cases where the LOG-a-TEC testbed has been used for performing experiments with cognitive radio.

The chapter is organized as follows. Section 13.2 discusses radio front-end for embedded testbeds and gives details for the ones used in LOG-a-TEC. Section 13.3 discusses the infrastructure of the testbed including deployment considerations, network design constraints, planning of the management network and the mechanisms that enable access, control and reconfiguration. Section 13.4 presents a signal generation use case in which the SNE-ISMTV-TI868 transceiver is used to emulate wireless microphone profiles. Section 13.5 presents an interference mitigation use case using a simple power allocation experiment on LOG-a-TEC. Finally, Sect. 13.6 summarizes the chapter.

13.2   Radio Front-Ends for Embedded Testbeds

Software defined radio has become an indispensable tool in communications research. It has allowed for very short development cycles for various protocols from the physical layer up. It has also significantly lowered the entry barrier to prototyping. With a software defined radio architecture a researcher only needs a personal computer with a relatively inexpensive radio frontend, while previously a fully equipped electronics laboratory was needed. The developer also does not need to know details of radio frequency electronic circuit design. In fact, physical radio front-ends can often be interchanged in a modern SDR design without much work thanks to layers of software that provide high-level abstractions over the hardware details. With the advent of frameworks like GNU Radio [6] and IRIS [7] one can develop complete RF systems in high level languages such as Python and XML or using user-friendly graphical tools like the GNU Radio Companion.

However this flexibility of SDR architectures and the convenience it offers to the developer comes with a cost in terms of processing performance, RAM and power requirements. While the throughput of embedded general-purpose processors is continuously improving, low-powered embedded devices are still a long way from supporting full-fledged SDR frameworks. One of the reasons is that battery capacity has not been scaling together with processor performance. It is likely therefore that battery powered and low cost devices will not be able to support the SDR development model for the foreseeable future.

On the other hand, the basic requirements for a cognitive radio can also be satisfied with architectures that are not fully software defined. There are commercially available low-power reconfigurable integrated transceivers that target developers of proprietary radio protocols for embedded devices. These are capable of covering a large part of radio requirements for simple cognitive terminal and dynamic spectrum
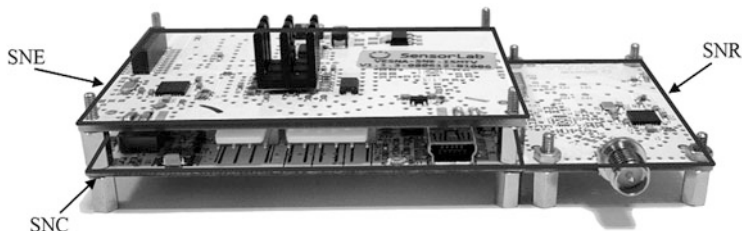
Fig. 13.1 Modular design of the VESNA wireless sensor network platform consisting of three types of modules, core (SNC), radio (SNR) and expansion (SNE)

access implementations: they are reasonably accurate at sensing the spectrum usage using energy detection, they have frequency agile local oscillators that permit fast frequency hopping and interference avoidance and allow software to reconfigure parameters of their transmissions like modulation scheme and channel bandwidth.

This chapter is focusing on cognitive radio testbeds based on low-power embedded devices and as a concrete representative of such devices we mostly refer to the VESNA sensor node platform [2].

13.2.1 VESNA with the SNE-ISMTV Expansion

VESNA [2] is a modular wireless sensor node platform generally consisting of three printed circuit boards: Sensor Node Core (SNC), Sensor Node Radio (SNR) and Sensor Node Expansion (SNE) as depicted in Fig. 13.1.

The SNC board contains a 32-bit ARM Cortex-M3 microcontroller running at up to 72 MHz CPU clock. It has 96 kB of RAM and 1 MB of flash program memory. Additional non-volatile memory is provided by a micro SD card. The microcontroller integrates 3 A/D converters that can be used to sample analog signals with up to 1 Msample/s. Two of these converters can also be combined to sample a single signal with 2 Msamples/s. Power is provided by a versatile subsystem that is capable of powering the sensor node from diverse sources, including photovoltaic cells and batteries.

VESNA supports a number of different wireless interfaces on the SNR board. Some SNR boards contain bare integrated transceivers while other carry radio modules that include both a transceiver and an additional microcontroller for network stack offloading. In the most common wireless sensor network deployment use case, VESNA nodes use a proprietary Atmel ZigBit low-powered radio module to connect to a wireless mesh network. In such use case, the SNE board is used for application specific hardware, usually to connect sensors to the node that cannot be accommodated by the interfaces on the core board itself.
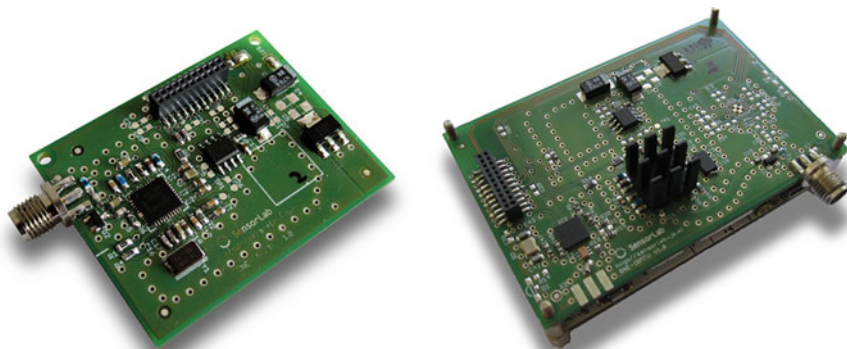
Fig. 13.2 UHF receiver expansion boards used for cognitive radio experimentation: early prototype (*left*) and final VESNA SNE-ISMTV-UHF circuit board (*right*)

For experiments with cognitive radio and deployment in cognitive radio embedded testbeds such as LOG-a-TEC, this setup has been left basically unmodified. However, instead of using the radio hardware on the SNR boards for experimental use, we have developed a series of SNE expansion boards that contain a separate set of transceivers. This had an important benefit of leaving the original mesh network interface on VESNA unmodified. That interface can then be used for a reliable wireless testbed management network, independent of the experimental work.

The series of SNE boards for cognitive radio experimentation have been named SNE-ISMTV [8], mnemonically identifying the supported frequency bands. Different versions of the SNE-ISMTV share a single printed circuit board layout. They only differ in electronic component placement. The SNE-ISMTV-TI2400 contains a Texas Instruments CC2500 transceiver for the 2.4 GHz band, SNE-ISMTV-TI868 a Texas Instruments CC1101 sub 1-GHz transceiver and SNE-ISMTV-UHF a custom designed energy detection receiver for the UHF band as shown in Fig. 13.2.

13.2.2 Reconfigurable Integrated Transceivers

A list of transceivers for sub-1 GHz frequencies that are commonly found on small embedded devices, such as wireless sensor nodes, is given in Table 13.1. Same vendors often offer similar devices that work in the 2.4 GHz international ISM band. While they are mostly controlled via a digital bus like SPI (Serial Peripheral Interface) or I^{2}C (Inter-Integrated Circuit) from a separate microcontroller on the embedded device, it is also common to find such devices integrated with a general-purpose processing core into a single chip.

A typical block diagram of these reconfigurable transceivers is shown in Fig. 13.3. They all have a single chip design that requires very little in terms of

Table 13.1 List of common reconfigurable sub-1 GHz transceivers

Transceiver	Modems	Bitrate (kB/s)	f_c (MHz)	BW (kHz)	Hop time (μs)
TI CC1101	FSK, GFSK, OOK ASK, MSK	0.6–500	779–928	58–812	75
Silicon Labs Si4313	FSK, GFSK, OOK	0.2–128	240–960	2.6–620	200
Nordic Semi nRF905	GFSK	50	430–928	100	650
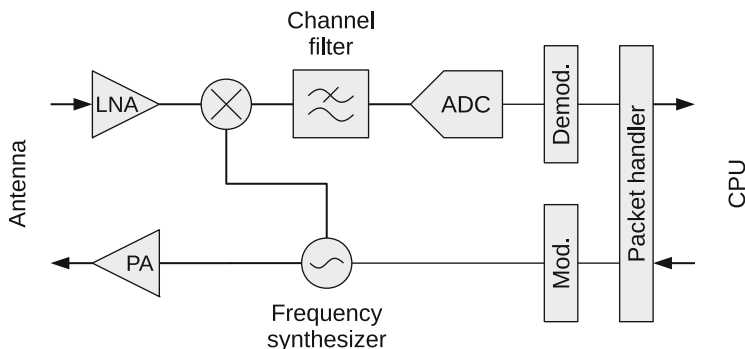
Fig. 13.3 Block diagram of a typical reconfigurable transceiver

external components. Rarely, they are coupled to an external power amplifier for range extension and/or an external low noise amplifier for improved noise figure.

A PLL frequency synthesizer is used both as a local oscillator in the reception mode and as a frequency agile sine-wave generator in transmission mode. A programmable digital or an analog filter bank provides channel selectivity that is reconfigurable from software. Several modulation and demodulation blocks can also be software selected. Additionally, packet handling logic can be enabled, which can be used to off-load start symbol detection, whitening, CRC calculation and similar tasks from the main CPU. Since the frequency synthesizer is shared between receive and transmit chains they typically allow only half-duplex operation.

Disabling packet handling hardware allows for continuous reception or transmission by streaming digital data to or from the transceiver. This allows for interpretation of such a stream as a sampled analog signal which opens interesting possibilities, like emulation of analog transmissions. It should be noted though that these transceivers do not typically allow access to the raw signal samples from the ADC. The only data stream available to the CPU is one provided by integrated demodulator blocks. This prevents development of spectrum sensing methods beyond energy detection and signal processing using such hardware.

Fig. 13.4 Typical flowchart for a procedure performing energy detection spectrum sensing on a part of spectrum between two RF channels, *start_chan* and *stop_chan*, with a reconfigurable transceiver

```
                    ( start )
                        │
                        ▼
            ┌───────────────────────┐
            │  current_chan =       │
            │  start_chan           │
            └───────────────────────┘
                        │
                        ▼
                ◇ current_chan >        true
                  stop_chan    ──────────►  ( stop )
                        │
                        │ false
                        ▼
            ┌───────────────────────┐
            │  tune PLL to          │
            │  current_chan         │
            └───────────────────────┘
                        │
                        ▼
            ┌───────────────────────┐
            │  wait for PLL to      │
            │  settle               │
            └───────────────────────┘
                        │
                        ▼
            ┌───────────────────────┐
            │  trigger RSSI         │
            │  measurement          │
            └───────────────────────┘
                        │
                        ▼
            ┌───────────────────────┐
            │  wait for RSSI        │
            └───────────────────────┘
                        │
                        ▼
            ┌───────────────────────┐
            │  read out RSSI        │
            └───────────────────────┘
                        │
                        ▼
            ┌───────────────────────┐
            │  increment            │
            │  current_chan         │
            └───────────────────────┘
```
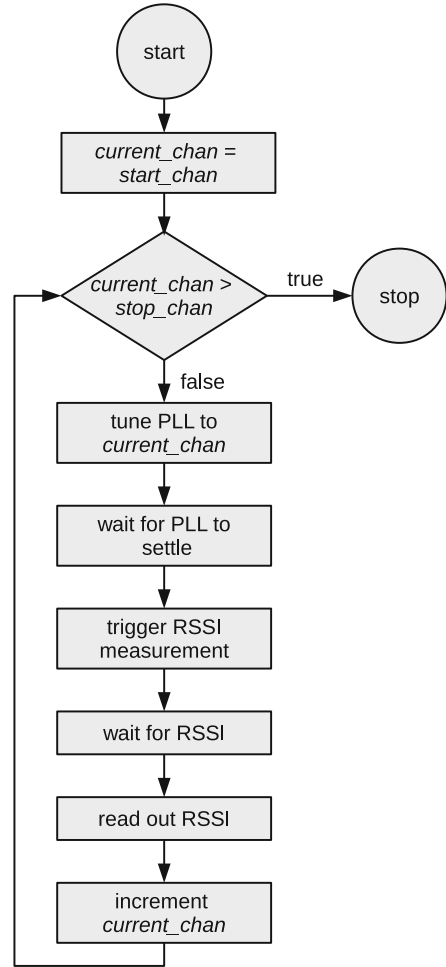
13.2.2.1 Energy Detection

Most transceivers include an integrated logarithmic detector behind the channel filter that can be used as a *received signal strength indicator (RSSI)*. While this capability is meant to be used for the implementation of clear-channel assessment, listen-before-talk and CS-MAC protocols, it can also be used as a simple energy detector for signals within the tuned RF channel. This capability is well known in the research community since it is often used for spectrum sensing experiments or as a low-cost spectrum analyzer. In this case, the transceiver is set to receive mode and swept through a list of channels as presented in the flowchart in Fig. 13.4. On each channel, a RSSI measurement is taken. Such measurements are limited in the resolution bandwidth by the available settings for the channel filter and in the sweep

time by the settling of various automatic gain control (AGC) stages in the receiver. Typical measurements take between 5 and 50 ms per channel and have a resolution of 0.5 dB.

When using the RSSI indicator in such a way, care must be taken with respect to the interpretation of the obtained results. Chips often have offsets that require calibration, if absolute values of the incident power are required. Detectors commonly also have a strong temperature dependency that might lead to wrong conclusions (as noted for example in [9]).

It is also important to note that since the logarithmic detector measures total power after the channel filter, its readings are only approximately equal to the signal power when signal power is significantly above the receiver noise floor. In cases where the signal to noise ratio is close or below 1, only reading the RSSI is not an accurate indicator of the signal level and more advanced methods are required, such as using link quality indicator (LQI) as indicator of signal to noise ratio [10].

13.2.2.2 Programming Interface

The connection between a low-powered transceiver and the central CPU on the embedded device usually takes the form of an embedded serial control bus, like SPI or I²C. The transceiver exposes a small address space (typically 8-bit) to the controlling CPU over this bus. Most transceivers also provide additional interrupt lines. These can be used to simplify software on the CPU and also provide a more accurate timing of packet reception or transmission than what can be achieved using a serial bus.

The radio configuration typically requires setting around 50 hardware registers accessible from the control bus. For example, the channel filter bandwidth may be set by writing appropriate values to bits 4 through 7 into an 8-bit register at address 0x10 as presented in Fig. 13.5 line 9. Usually the exact register settings are derived from the calibration values that are only known to the transceiver manufacturer. Because of that, the specified performance of the radio can often only be achieved by using values calculated by specialized software for each transceiver configuration such as the one depicted in Fig. 13.6. In such case, a cognitive terminal must store a predefined set of radio configurations that have been precalculated using the manufacturer's software.

Data reception and transmission is similarly performed through FIFO buffers accessible over the control bus, although sometimes a completely separate digital bus is used for streaming data as is the case for example in CC1101 in continuous transmission/reception mode.

There is currently no unified interface that would allow for fast prototyping of solutions using such hardware. Compared to high level scripting provided by the GNU Radio platform, here typically the developer has to have intimate knowledge of the radio hardware. In the future, standardised transceiver APIs might improve this situation. Similarly, there has been recently some development in open source SDR frameworks that allow for delegation of certain signal processing tasks to

```
1.  uint8_t initSeq[] = {
2.      CC1100_IOCFG2,   0x0B, // GDO2Output Pin Configuration
3.      CC1100_IOCFG0,   0x0C, // GDO0Output Pin Configuration
4.      CC1100_PKTCTRL0, 0x22, // Packet Automation Control
5.      CC1100_FSCTRL1,  0x08, // Frequency Synthesizer Control
6.      CC1100_FREQ2,    0x58, // Frequency Control Word
7.      CC1100_FREQ1,    0xE3, // Frequency Control Word
8.      CC1100_FREQ0,    0x8E, // Frequency Control Word
9.      CC1100_MDMCFG4,  0x86, // Modem Configuration
10.     CC1100_MDMCFG3,  0x75, // Modem Configuration
11.     CC1100_MDMCFG2,  0x00, // Modem Configuration
12.     CC1100_MDMCFG0,  0xE5, // Modem Configuration
13.     CC1100_DEVIATN,  0x67, // Modem Deviation Setting
14.     CC1100_MCSM0,    0x18, // Main Radio Control State
15.                            // Machine Configuration
16.     CC1100_FOCCFG,   0x16, // Frequency Offset Compensation
17.     CC1100_FSCAL1,   0x00, // Frequency Synthesizer Calibration
18.     CC1100_FSCAL0,   0x11, // Frequency Synthesizer Calibration
19.     CC1100_PATABLE,  0xFE, // TX power 0 dBm
20.     0xFF,     0xFF
21. };
22.
23. static void radioSetup(void)
24. {
25.     vsnCC_disableRadioInterrupt();
26.     vsnCC_init();
27.     vsnCC_radioReset();
28.
29.     vsnCC_strobe(CC1100_SIDLE);
30.     radioWaitState(CC_MARCSTATE_IDLE);
31.
32.     int n;
33.     for(n = 0; initSeq[n] != 0xff; n += 2) {
34.         uint8_t reg = initSeq[n];
35.         uint8_t value = initSeq[n+1];
36.         vsnCC_write(reg, value);
37.     }
38. }
```

Fig. 13.5 Example code required to setup a CC-series transceiver for continuous transmission on VESNA platform

hardware blocks like the ones found on the reconfigurable transceivers discussed here (for example specialized engines in IRIS). However, these frameworks still remain much too large for a typical microcontroller system today.

13.2.2.3   SNE-ISMTV-TI868 and SNE-ISMTV-TI2400

As an example of using low-power embedded devices for energy detection based spectrum sensing we refer in the following to two versions of purposely developed SNE-ISMTV expansion board for VESNA using Texas Instruments CCxxxx transceivers with the main specifications summarised in Table 13.2 for SNE-ISMTV-TI868 and in Table 13.3 for SNE-ISMTV-TI2400. SNE-ISMTV-TI868 uses the CC1101 chip and operates in the European 868 MHz short range devices (SRD) unlicensed band and upper channels of the UHF broadcast band. SNE-ISMTV-TI2400, on the other hand, uses the CC2500 chip and operates in the 2.4 GHz unlicensed international industrial, scientific, medical (ISM) band. Except for the difference in the radio frequency, which requires different antenna matching circuit, the two boards are mostly identical in hardware design. Similarly, the software interface of CC1101 and CC2500 is largely identical, which greatly simplified driver software.
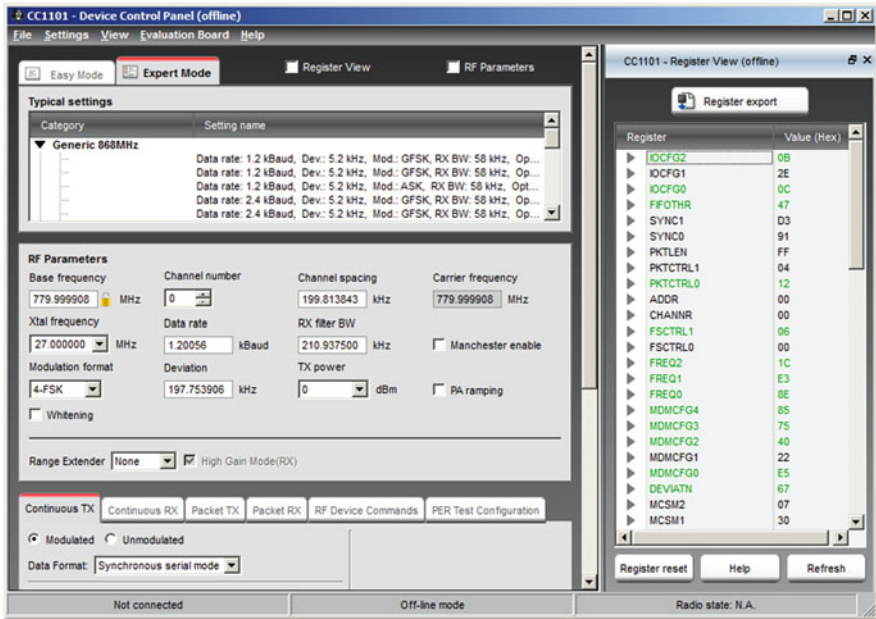
Fig. 13.6 Screenshot of Texas Instruments SmartRF studio software

Table 13.2
SNE-ISMTV-TI868
specifications

Parameter	Value
Central frequency range	780–871 MHz
Central frequency resolution	50 kHz
Channel filter bandwidth	60, 100, 200, 400 or 800 kHz
Power detector range	-123 to 4 dBm
Power detector resolution	0.5 dBm
Average noise level (1 Hz BW, 868 MHz)	-150 dBm
Channel sampling time	5 ms
Transmission power range	-30 to 12 dBm
Transmission power resolution	2 dBm

The analogue receive path on CCxxxx series transceivers consists of a low-noise amplifier, quadrature down-converter and a channel filter. Additional filtering, gain control and demodulation are performed in the digital domain, after the signal has been digitized by an analogue-to-digital converter (ADC) at an intermediate frequency. A fractional-N PLL frequency synthesizer is used as a local oscillator. For transmission, the same frequency synthesizer is used to produce a modulated quadrature signal which is amplified by a power amplifier stage. 50 Ω antenna matching is performed by an external LC balun network.

The transceiver is connected to the sensor node core CPU with an SPI bus, which exposes radio configuration registers and FIFO buffers. This low-level interface to

Table 13.3
SNE-ISMTV-TI2400
specifications

Parameter	Value
Central frequency range	2400–2480 MHz
Central frequency resolution	50 kHz
Channel filter bandwidth	60, 100, 200, 400 or 800 kHz
Power detector range	−123 to 4 dBm
Power detector resolution	0.5 dBm
Average noise level	−159 dBm
(1 Hz BW, 2400 MHz)	
Channel sampling time	5 ms
Transmission power range	−55 to 1 dBm
Transmission power resolution	2 dBm

the transceiver has a form of a finite state machine. Digital radio control logic allows reconfiguration of the frequency synthesizer settings (base frequency, channel spacing) and baseband channel filter bandwidth. A logarithmic-response detector can be used to measure the RSSI. Baseband modulator and demodulator blocks are capable of 2-FSK, 4-FSK, GFSK, MSK and ASK/OOK modulations. Continuous transmission and reception using these modulations is possible with data streamed via a dedicated digital bus on the transceiver's auxiliary pins. Also available is a pseudo-random sequence generator which can be used in continuous transmission mode.

For packet-based transmissions, integrated packet handling hardware implements a proprietary packet encapsulation scheme. This makes it mostly incompatible with standard MAC layers like the one defined by IEEE 802.15.4. The hardware is capable of preamble and synchronization word detection, data integrity check using CRC, address filtering and data whitening.

In cognitive radio experiments so far, SNE-ISMTV-TI868 and SNE-ISMTV-TI2400 have been used as narrow-band energy detection spectrum sensing receivers using the RSSI functionality in the 868 MHz and 2.4 GHz bands to build radio environment maps or to supply channel occupancy information to cognitive terminals.

VESNA nodes with SNE-ISMTV expansion have also been used as controlled interferers and targets for spectrum sensing using other devices. In this case, the continuous transmission mode of the transceiver has been used. The data sent was either a pseudo-random sequence generated by the transceiver or a stream generated on the sensor node's CPU. With the latter option we have been able to approximate analogue transmissions generated by wireless microphones.

To help with the problem of complicated low-level interface exposed by this type of hardware an abstraction layer for software running on the sensor node has been developed. This simplifies programming for two common tasks: energy detection and signal generation. For each version of the SNE-ISMTV board, several hardware profiles have been defined that specify PLL settings (channel base and spacing) and modem configuration.

13.2.3 Getting Closer to SDR with Custom Hardware

Microcontrollers that integrate multiple analog-to-digital and digital-to-analog converters in the Msample/s range make for an inviting target for an embedded version of a software-defined architecture. With a suitable front-end and anti-aliasing filters these can be used to generate or sample signals directly.

However, the processor throughput of such devices is typically far behind the capability of processing a continuous stream from its ADC in real-time. For example, a STM32F1 family microcontroller running at 64 MHz, which is near the top of the range for an ARM Cortex M3 core without DSP extensions, will only process signals on the order of 100 ksample/s even when doing the simplest signal processing tasks.

Processors with large RAM sizes have the possibility of capturing an interval of signal samples and then doing off-line processing. This type of processing, however, has limited usability in cognitive radio applications beyond spectrum sensing.

#### 13.2.3.1 TV Band Spectrum Sensing with SNE-ISMTV-UHF

The VESNA expansion board for spectrum sensing in the UHF and VHF bands has been developed specifically for research into heterogeneous spectrum sharing and cognitive radio applications in the TV-band whitespaces. It is based on a cheap and compact multi-standard TV silicon tuner design which allows it to be deployed in large sensor networks for distributed sensing applications.

SNE-ISMTV-UHF contains a VHF and UHF TV band receiver based on the NXP TDA18219HN silicon tuner and was designed for spectrum sensing experimentation in TV white spaces (TVWS). Figure 13.2 presented an early version of the receiver and the final printed circuit board used in LOG-a-TEC. The TDA18219HN silicon tuner integrates a low-noise amplifier, RF tracking filters, single down-conversion low intermediate-frequency image-rejection mixer, frequency synthesizer and selectable channel filters. It also includes multiple stages of analogue automatic gain control (AGC).

Individual receiver stages can be reconfigured through a state machine with an I^{2}C bus interface. Digital control logic also controls built-in test tone generator and RF filter calibration.

For energy detection, SNE-ISMTV-UHF offers two detectors with a logarithmic response. A high-range detector is built into TDA18219HN and has a range from −80 to 0 dBm with 1 dBm resolution. The measurement is controlled via a state machine internal to TDA18219HN which is armed and triggered through the digital I^{2}C bus. The measurement process includes signal averaging.

For measuring signals below −80 dBm, a demodulating logarithmic amplifier Analog Devices AD8307 is connected between the TDA18219HN intermediate frequency output and the 1 Msample/s analogue-to-digital converter on the VESNA SNC (Sensor Node Core) as depicted in Fig. 13.7. This puts sampling of the signal power level in control of the firmware, allowing various averaging and sampling
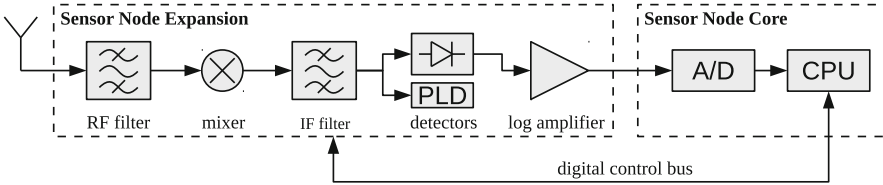
Fig. 13.7 Block diagram for signal reception using VESNA with SNE-ISMTV-UHF

methods. Additionally, gain in the intermediate frequency stage can be adjusted via a variable-gain amplifier controlled by a digital-to-analogue converter (DAC) in the SNC.

To lower the power consumption, both the logarithmic amplifier (via ENB pin) and the tuner (via I^2C sleep-mode control registers) can be powered down.

Two hardware profiles are provided that differ in energy detector resolution bandwidth. A wide-band setting sets the channel filter to 8 MHz, allowing energy detection at the bandwidth of a complete DVB-T channel. A narrow-band setting sets the channel filter to 1.7 MHz and is suitable for detection of wireless microphones and secondary users in the TV white spaces.

While TDA18219HN and the SNE-ISMTV designs allow operation down to 42 MHz, there are currently no hardware profiles available for VHF frequency band.

Since it is based on a DVB-T tuner, this extension board is not capable of signal transmission in contrast to the extension boards SNE-ISMTV-TI868 and SNE-ISMTV-TI2400, as described in the previous section.

13.2.4   Summary

In summary, reconfigurable low-power transceivers are capable of performing many radio frontend duties in a cognitive radio setup. However the development of solutions is far from being as accessible as with software defined radio architectures. Such devices deployed in a testbed can be used for instance in a supporting role for other, more full featured cognitive terminals. They can perform energy detection spectrum sensing in a distributed fashion. Reconfigurable transceivers are also flexible enough to take an active role in certain dynamic spectrum access schemes.

With SNE-ISMTV-UHF, we are looking at ways of getting closer to software defined radio using a custom designed low-cost receiver for the VHF and UHF bands based on a commercial integrated silicon tuner. In contrast to reconfigurable transceivers, however, devices with tuners can only be used in cognitive radio applications for spectrum sensing, with capabilities depending on the accessibility of the signal at different stages within the receiver.

13.3 Testbed Infrastructure

In order to perform real-life experiments with communication protocols, representative testbeds are needed. Such testbeds can typically consist of a few devices up to hundreds of devices. On the devices experimental implementations of protocols are being run in various scenarios and the results are recorded and later analyzed. This way the performance of the protocols can be evaluated and possibly improved. To facilitate the research activity in dynamic spectrum allocation we created a testbed called LOG-a-TEC primarily for experimentation with spectrum sensing, cognitive radio networking and testing of arbitrary protocols on low-cost low-complexity hardware.

In order to use the embedded devices such as VESNA based sensor nodes with expansion boards described in Sect. 13.2 in a cognitive radio testbed, the overall testbed infrastructure needs to be carefully planned and set up accordingly, including (i) the management network, wireless or wired, for configuration of nodes and collection of sensing data, (ii) locations for mounting nodes and antennas, (iii) provision of power for nodes in case of external powering, (iv) adequate data storage and processing capabilities, and (v) a suitable system administration user interface.

Because a cognitive radio testbed is used to experiment with the communication between the nodes, the radio interface used by experiments is unreliable and nodes need to be equipped with additional radio interface for the management network to provide remote access to the nodes.

A testbed infrastructure may also provide power to the nodes in the testbed. In comparison to conventional testbeds, where external power is a must, the physical requirements for the infrastructure are much more relaxed in embedded testbeds. Nodes can be battery powered for the duration of the experiment. Also the management network need not to consist of gigabit Ethernet cabling but can be wireless.

While an embedded testbed is focused on cognitive radio applications that might one day be used in mobile, battery powered devices, most such testbeds are not themselves battery powered, especially those that are permanently deployed. A testbed might use low-powered processors and radio to conduct exclusively tests of wireless protocols and does not focus on the power consumption aspect of its operation. Thus it is not uncommon that in an embedded testbed each low-powered node is accompanied by a conventional, networked computer that provides management functionality. This significantly simplifies management of the testbed. Embedded operating systems running on nodes often do not support run-time reprogramming or include system administration functionality that is common on multi-user systems like GNU/Linux.

13.3.1 LOG-a-TEC Testbed Deployment

The LOG-a-TEC testbed was deployed in the town of Logatec, Slovenia [11, 12]. It consists of two clusters, each cluster being composed of 25 devices. One cluster
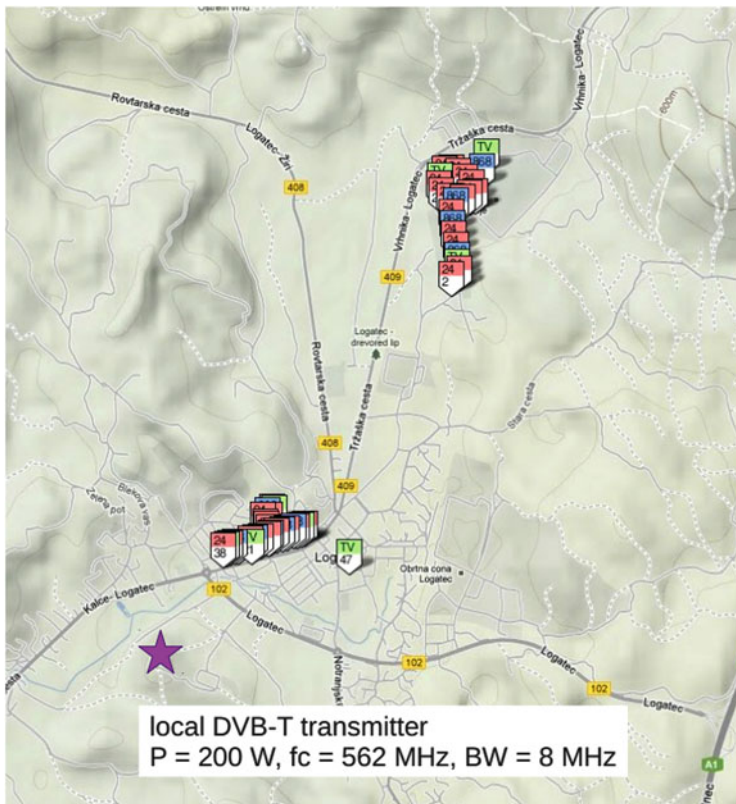
Fig. 13.8 Locations of the device clusters inside Logatec

is located in the industrial zone and the other in the town center. The location of the two clusters inside Logatec is presented in Fig. 13.8. The pentagonal markers represent the locations the nodes. The cluster of markers situated in the upper side of Fig. 13.8 corresponds to the industrial zone, while the cluster of markers on the lower side of the same figure corresponds to the town centre. The majority of devices in both clusters are installed on public light poles and connected to external power supply.

The deployed devices are based on the VESNA hardware platform [2] described in Sect. 13.2. Each VESNA is equipped with multiple radio modules:

- Atmel's ZigBit radio module ATZB-900-B0 operating in the license-free 868 MHz European SRD band is used for the management network.
- One of the three radio modules for spectrum sensing and cognitive radio experimentation that can be hosted on the SNE-ISMTV extension board described in Sect. 13.2.1, i.e. CC1101 and CC2500 from Texas Instruments or the custom designed spectrum sensing receiver operating in the UHF TV band.
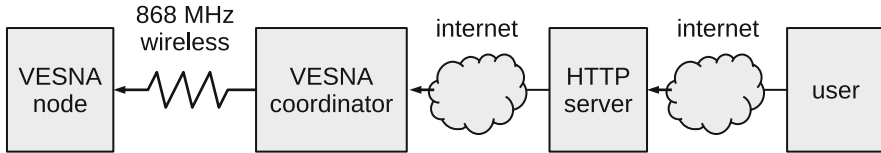
Fig. 13.9 The logical block diagram of one cluster in LOG-a-TEC

- Atmel's IEEE 802.15.4 compliant radio AT86RF212 operating in the license-free 868 MHz European SRD band for communication experiments.

Each VESNA is thus equipped with one of the following three combinations of the radios:

- ATZB-900-B0 + CC1101 for 868 MHz sensing + AT86RF212
- ATZB-900-B0 + CC2500 for 2.4 GHz sensing + AT86RF212
- ATZB-900-B0 + TV receiver for UHF sensing + AT86RF212

The logical block diagram of one cluster of the testbed is depicted in Fig. 13.9. The cluster has two parts that are situated at different physical locations. The first part, where the experiments will be performed, is located in Logatec and is illustrated on the left side of Fig. 13.9 and consists of a set of VESNA nodes and the cluster coordinator. This part is connected to the Internet through its Gateway that corresponds to the VESNA coordinator and internally uses a ZigBee network for communication. The second part of the testbed, presented on the right hand side of Fig. 13.9, is the infrastructure serving the experimental part. This part consists of a PC running an HTTP server that is connected to the Internet and communicating with the Gateway. This server provides an HTTP API which supports programmatic access to VESNA's REST interface and also hosts the LOG-a-TEC web portal depicted in Fig. 13.10. This portal renders a map with the location of VESNAs and their status, allows manually issuing GET and POST requests and supports running simulations with the GRASS-RaPlaT [13] tool.

The currently supported VESNA software stack is depicted in Fig. 13.11 and can use libopencm3 or STM32FWLib as an abstraction layer towards hardware registers and MCU peripherals. While the first library is released under an open source license (LGPL), the second is proprietary. VESNA drivers then provide a high level abstraction to support storage, various sensors and functionalities of the platform which may run the event driven Contiki operating system if needed. In many applications, including LOG-a-TEC, we use application specific software without an operating system. On VESNA, an application can run (1) on top of the Contiki OS, (2) run without an OS using the drivers or (3) run without an OS using libopencm3. The decision on how to build an application depends on licensing constraints and on the complexity of the application. For the case of LOG-a-TEC, the second option using STM32FWLib, VESNA drivers and a custom application was selected. The application provides (1) a REST interface that supports remote
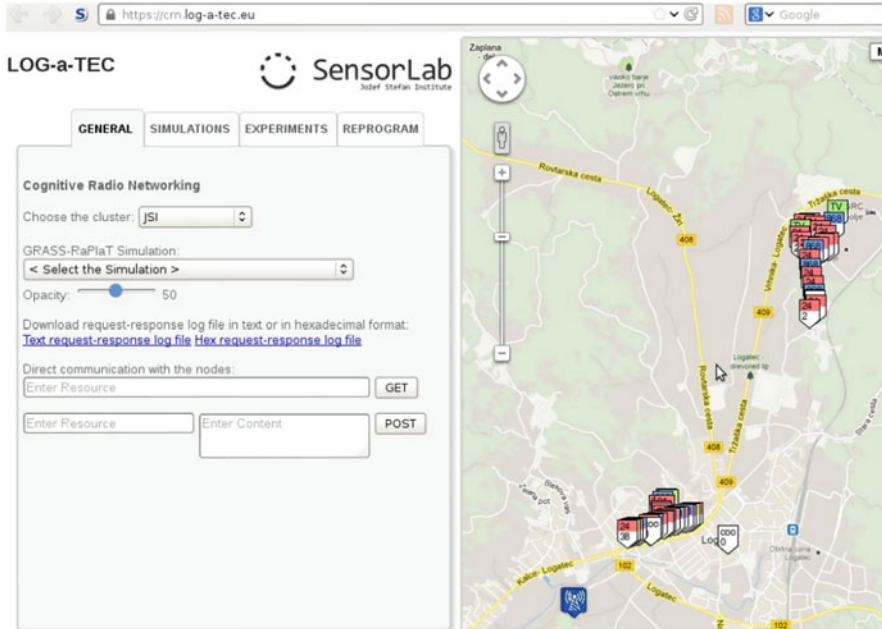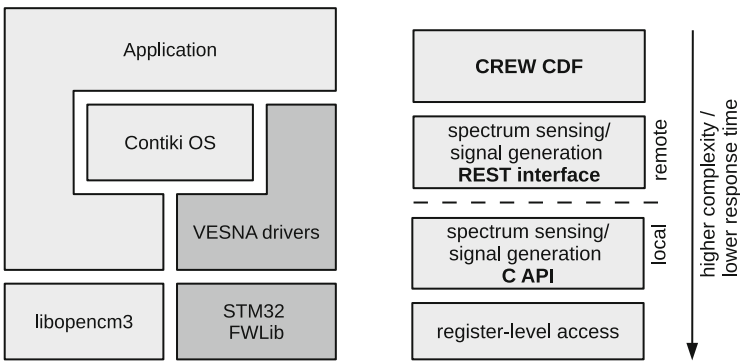
Fig. 13.10 The LOG-a-TEC web portal

Fig. 13.11 VESNA software stack (*left*) and layers of abstraction over SNE-ISMTV expansion boards (*right*)

experiments, (2) native code for local experiments and (3) testbed management functionalities such as over-the-air reprogramming and monitoring.

The control of the SNE-ISMTV extension for the purpose of experimentation can be achieved in several ways as illustrated in Fig. 13.11. Currently there are four ways of controlling an experiment that differ in terms of complexity, response time and location. The higher the complexity required to control the hardware to perform

the experiment, the lower the response time. Additionally, there are two options that allow remote control of the experiment and two options that allow local only control requiring physical connection to the hardware.

A low-complexity approach that supports remote experiment control using a device and testbed independent language using the XML syntax is to define an experiment using the CREW common data format (CDF) and then request the LOG-a-TEC infrastructure to run the defined experiment. The resulting measurement will be provided in the form of CSV or Matlab files containing time, frequency and power values resulting from the sensing experiment. A slightly higher complexity approach that also supports remote experiment control is to use the available REST interface. This assumes the experimenter downloads the available Python or Javascript tools[1] and writes the logic of the experiment using these tools. Using these tools, the experimenter is able to specify actions such as (i) sense band B at time T using device D and method M, and (ii) transmit at frequency F and power P. This second approach is the most frequently used.

A higher complexity approach to controlling the experiment is via direct access to the C API from native code. This option removes network latency caused by network round trips, on-board processing and interference from the wireless management network operating on 868 MHz. Finally, the most complex and time consuming way of controlling the experiment is to directly program the hardware (i.e. the CC1101, CC2500 and TDA18219 chips) and exploit the full capabilities of the hardware. This approach is time consuming as it requires testing but results in lower response time in terms of spectrum sensing and transmission.

13.3.2   Network Design Constraints

The design of ZigBee based wireless management network, which provides remote operation of the deployed testbed, needs to take into consideration the constraints of the devices and possible deployment locations [14]. One of the constraints of the design of the network is the availability of frequency bands where the ZigBee network can operate. There are two choices for this network, the 2.4 GHz ISM band and the 868 MHz SRD band. The 868 MHz SRD band offers better propagation characteristics and it does not interfere with the crowded 2.4 GHz ISM band, where majority of spectrum sharing experiments are expected to be performed as opposed to TVWS related experiments in UHF band. Thus, the 868 MHz SRD band was chosen as the operating band for the management network in the LOG-a-TEC testbed.

The locations for devices are constrained by the following factors:

- Availability of Internet access: the Gateway of each cluster has to be connected to the Internet; for maximizing the performance, wired connection is preferred.

---

[1] The tools are available for download at https://github.com/sensorlab

- Location of light poles (or alternative infrastructure): devices are mounted on light poles belonging to the public lightning system. Possible exceptions are the Gateway devices.
- Power connections to the light poles: the devices in LOG-a-TEC have always available power source. Light poles are powered in groups, so economically the most suitable solution is to equip all light poles of the selected light pole groups.
- Connectivity: because all the devices will be accessed through the Gateway, for maximum performance it is important to have minimum number of hops from any device to the Gateway.
- Possibilities for experiments: the clusters of the testbed should allow experiments that involve different types of network topologies, including multi-hop scenarios and mesh networks.

The first two constraints drastically reduce the number of possible locations for the devices. The problem of selecting the device locations for both clusters can be modeled by the abstract problem of selecting $N = 25$ locations from the available M possible locations. This model is used in later sections.

Besides constraints regarding network connectivity it is important to evaluate also the network performance required for the operation of the testbed. The usual management activity for the network generates small amount of traffic while the following activities require the transfer of large amounts of data:

- Transmission of firmware images for devices, in order to run custom applications during experiments: For testing custom protocols, their implementation has to be uploaded to the devices. Because a local storage for firmware images is implemented on each device, the uploading operation is not time-critical. The amount of transferred data is expected to be several hundreds of kilobytes for each different firmware type. Since the firmware distribution mechanism makes use of the broadcast nature of radio communications, multiple nodes can receive a firmware image instantly by a single transmission.
- Collection of spectrum sensing data: During spectrum sensing experiments significant amount of data is collected in the testbed, which can consist of received signal strength indicator (RSSI) values taken at a given frequency and at a given time, or it can consist of various performance metrics like packet loss rate or average network throughput. The spectrum sensing data can be similarly saved to a local storage on devices as the firmware images, but the resulting amount of data can be in the order of megabytes. If the required sample rates of the data allow it, real-time data collection can be implemented; in this case the measurement results are transmitted as soon as they are collected. The feasibility of the real-time data collection depends on the latency and throughput of the ZigBee based management network and on the performance of the sensing radios.

For the management network operating in the 868 MHz SRD band, the Atmel's ATZB-900-B0 ZigBit™modules [15] are used. The modules implement the ZigBee specification with proprietary extensions. When modules are started, they form a

mesh network with a central coordinator, zero or more routers and zero or more end devices. In both clusters of the testbed, the VESNA node with the Internet access is configured as the coordinator of the ZigBit network. All the remaining nodes in the testbed are configured as routers, since this configuration appears to support larger throughput compared to end devices, most likely because of radio duty-cycle limitation of the latter configuration.

The maximum amount of data that can be transmitted in a ZigBit frame is 94 bytes when security is disabled and 64 bytes when security is enabled [16]. Enabling security means that the traffic in the network is encrypted. This feature is useful for preventing various types of attacks on the testbed. Because no application level security is planned, enabling security in the ZigBit network is necessary for preventing malicious packet injection in the network.

Tests have shown that maximum 10 packets per second can be transmitted before significant packet loss happens in the ZigBit network. This means that the maximum point to point throughput in the network is $10 \times 64 = 640$ bytes/s. For calculations, this number can be used as an approximate measure of the network performance. For the transmission of firmware images, the performance of the network can be increased by broadcasting the firmware image in the network; in this case the redundant transmissions to each node can be eliminated.

13.3.3 Performance Evaluation of the Management Network

Before the actual deployment of the testbed the behavior of the management network was determined experimentally by measuring radio link quality between candidate locations of VESNA nodes. For the experiments, two VESNA nodes were used, both equipped with the ATZB-900-B0 radio modules. One VESNA was acting as an echoing device denoted with (E), while the other VESNA was transmitting requests to the first one and receiving the responses, hence it is denoted with (T/R). This setup is schematically depicted in Fig. 13.12. The VESNA node (E) was responding to the requests sent from the VESNA node (T/R). The VESNA node (T/R) was connected with an RS232 connection to a PC which was saving the measurement logs. The PC and the VESNA node (T/R) connected to it were moving during the measurement campaign between candidate locations, while the VESNA node (E) was fixed to selected location. Both nodes involved in measurements are depicted in Fig. 13.13.

The saved output from the VESNA node (T/R) consists of:

- Sequence number of the reply; this number allows detection of lost requests or other unexpected problems.
- Round trip time: the time elapsed between the moment of sending the request and the reception of the reply.
- Received Signal Strength Indicator (RSSI): the power level of the received reply, measured in dBm.
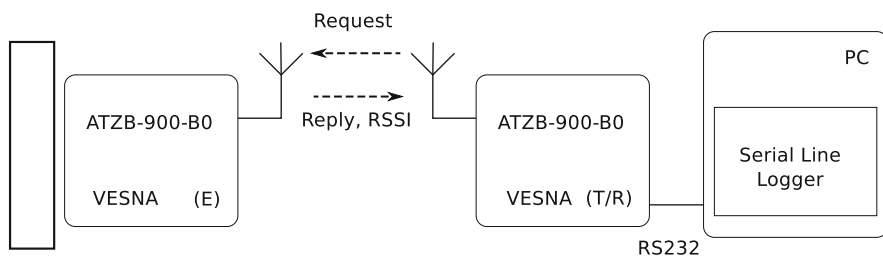
Fig. 13.12 Measurement setup

Fig. 13.13 VESNA node (T/R) with a PC on a mobile measurement platform (*left*) and VESNA node (E) at a fixed location (*right*)

- Link Quality Indicator (LQI): value showing the power and the amount of distortion in the frames received by the radio [15]. Because the ATZB-900-B0 modules are internally implemented with an AVR microcontroller (ATmega1281) and an AT86RF212 radio, the LQI obtained from the modules is expected to correspond to the LQI of the AT86RF212 radio, as presented in the radio's datasheet [17].

For every 30 requests, the VESNA node (T/R) printed the statistics about packet loss, time needed for measurements, average round trip time and RSSI. A representative part of the output from the VESNA node (T/R) is presented in Fig. 13.14.

```
reply from 5051: seq=266 time=108 ms
    lqi=252 rssi=-54 dbm
reply from 5051: seq=267 time=105 ms
    lqi=252 rssi=-52 dbm
reply from 5051: seq=268 time=106 ms
    lqi=252 rssi=-53 dbm
30 pkts transmited, 30 received,
    0% packet loss, time 30023 ms
    avg 109 ms rtt, -52 dbm rssi
```

Fig. 13.14 Statistics about packet loss provided for each 30 requests

This data was complemented by a measurement location record in terms of GPS coordinates obtained by a dedicated GPS receiver, a GPSMAP 60CS device produced by Garmin.

13.3.4 Measurements in Logatec

Based on the methodology for performance evaluation of management network, described in Sect. 13.3.3, field measurements have been carried out at candidate locations for deployment of VESNA based spectrum sensing nodes in two clusters in Logatec. The obtained measurement results are presented in the following subsections.

13.3.4.1 Measurement Results in the Industrial Zone Cluster

The locations of the light poles in the industrial zone are depicted in Fig. 13.15 on a satellite and on a map view of the zone. Figure 13.15 also shows the numbers assigned to the light pole locations. Location (C) represents the only location with available fixed Internet access, making it the most suitable candidate location for the coordinator.

Two sets of measurements have been carried out in this zone: in one set the link quality between the position of the coordinator and one subset of the light pole positions has been measured, and in the second set the link quality between location number 5 and another subset of the light pole locations has been measured. The two subsets of measurement locations partially overlap, and together they cover all possible device locations. The results from both first sets of measurements are presented in Fig. 13.16.

From the results it can be seen that for locations farther away from the coordinator than location 5, over 10 % of the transmitted packets is lost. Hence it has been considered that the nodes that are closer than location 5 are reachable with one hop, and the rest of the nodes need more than one hop to reach the coordinator. Because of the bad connection quality the second set of measurements has been carried out with the echoing device placed at location 5. The results from the second set of experiments show that most of the positions are reachable with two hops,
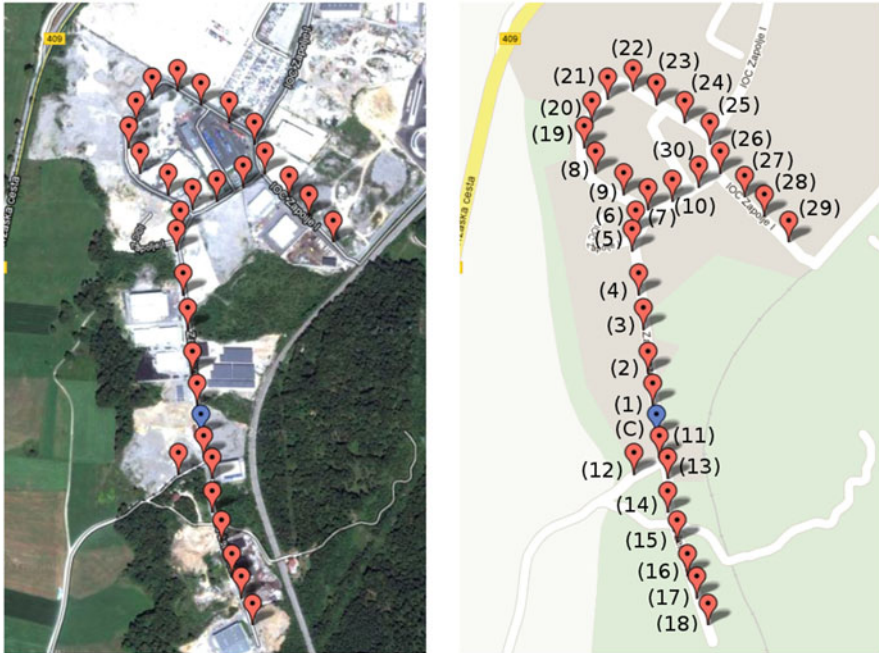
Fig. 13.15 Satellite (*left*) and map view (*right*) of measurement locations in the industrial zone of Logatec. Location designations are given in *parentheses*

although some locations, for instance 20, 25, 26, 28 and 29 might need a third hop for reaching the coordinator. From the measurements it has been concluded that the ZigBit modules operating in the 868 MHz SRD band can reach nodes at least at 5 light poles away. These estimates are pessimistic because the measurement have been carried out at approximately 1.2 m above the ground level, whereas nodes were to be installed on the public light poles at approximately 4–5 m above the ground level, where propagation conditions are expected to be more favourable.

13.3.4.2   Measurement Results in the City Center Cluster

The locations of the light poles in the center of Logatec are presented in Fig. 13.17. The numbers on the map represent the numbers given to the public light poles in the center of Logatec. The location marked with (C) shows the preferred candidate location for the coordinator, providing fixed Internet access. The public light poles in the center are installed more densely than in the industrial zone.

Similar as in the industrial zone, also in the center of Logatec two sets of measurements have been carried out. In one set the link quality from the location of the coordinator towards most of the other possible locations has been estimated, and
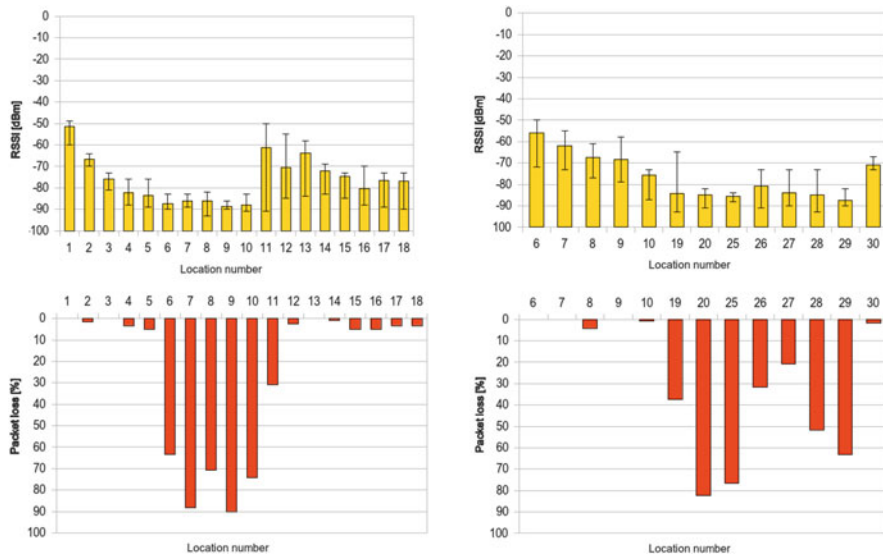
Fig. 13.16 RSSI and packet loss in the industrial zone measured between light poles and the coordinator's location (*top* and *bottom left*) and measured between light poles and location number 5 (*top* and *bottom right*)

in the second one the link from location number 25 towards the rest of the public light poles in the zone has been estimated.

The results of the two measurement sets are shown in Fig. 13.18. From the first set of measurements it can be seen that the coordinator has a good quality link to all of the devices installed along the central street of Logatec. The only location where more than 10 % packet loss has been measured is location 20. After performing the measurement of links pointing to the coordinator's location, the echoing device has been moved to location 25, from where good link quality was expected towards the rest of the possible locations of devices. The second set of results shown in Fig. 13.18 confirms the good link quality towards locations 26–34. Based on these measurements, all nodes from the cluster in the center of Logatec are expected to be at most two hops distance from the coordinator of the cluster.

### 13.3.4.3 Selecting Device Locations for Testbed Deployment

ZigBit modules used in the LOG-a-TEC testbed for the management network form a network with a tree topology starting from the coordinator as the root of the tree. Based on the measurement results given in Sects. 13.3.4.1 and 13.3.4.2 the maximum hop count in network when communicating with the gateway is expected to be 3 or 4 hops in the case of the industrial zone and 2 hops for the center of Logatec, so the proper functioning of the ZigBit modules and the automatic build-up of the ZigBit network is expected.
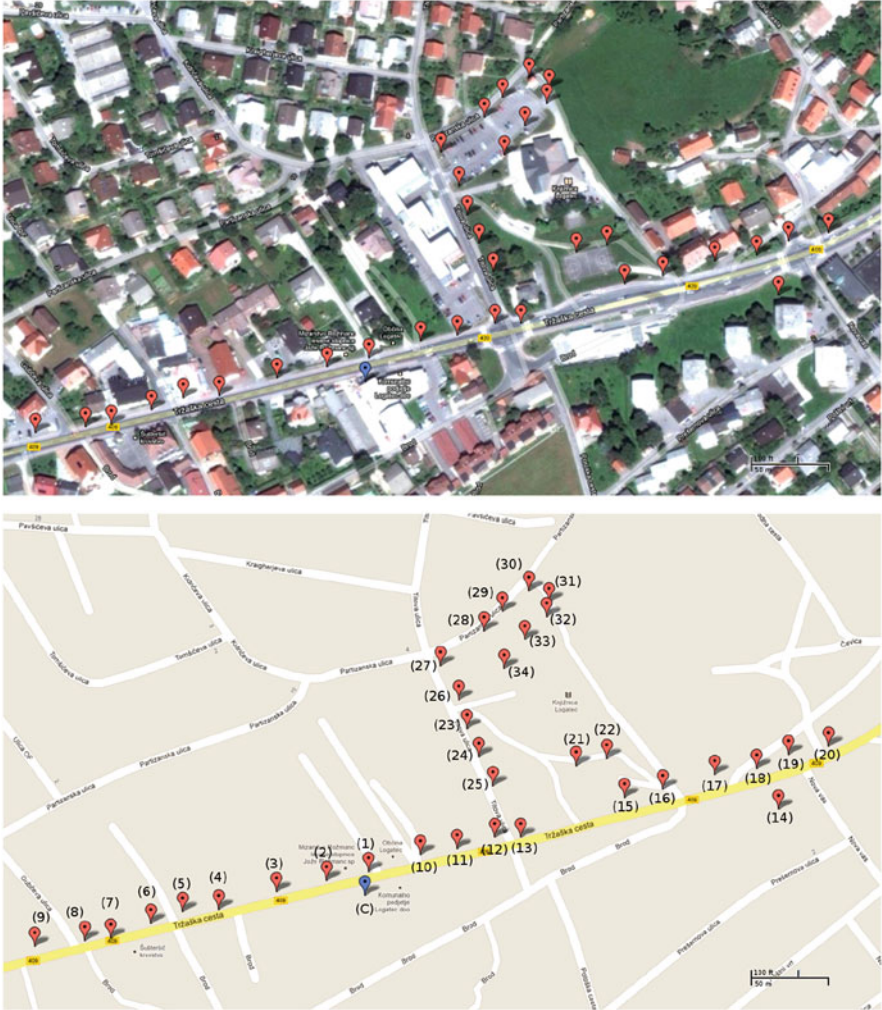
Fig. 13.17 Satellite (*top*) and map view (*bottom*) of measurement locations in the center of Logatec. Location designations are given in *parentheses*

From the measurements it can be seen that for the industrial zone cluster the communication range in the 868 MHz SRD frequency band is at least 5 light poles, and even larger for the city center with more dense public light poles. This means that VESNA nodes with SNE-ISMTV-TI868 operating in the 868 MHz SRD band can be installed on every third light pole, without affecting the sensing performance significantly. VESNA nodes with SNE-ISMTV-TI2400 operating in the 2.4 GHz ISM band have been installed every second light pole or closer, because they have smaller reception and transmission range. TV transmissions in the UHF frequency band are constant and the locations of the transmitter towers are known, so it is
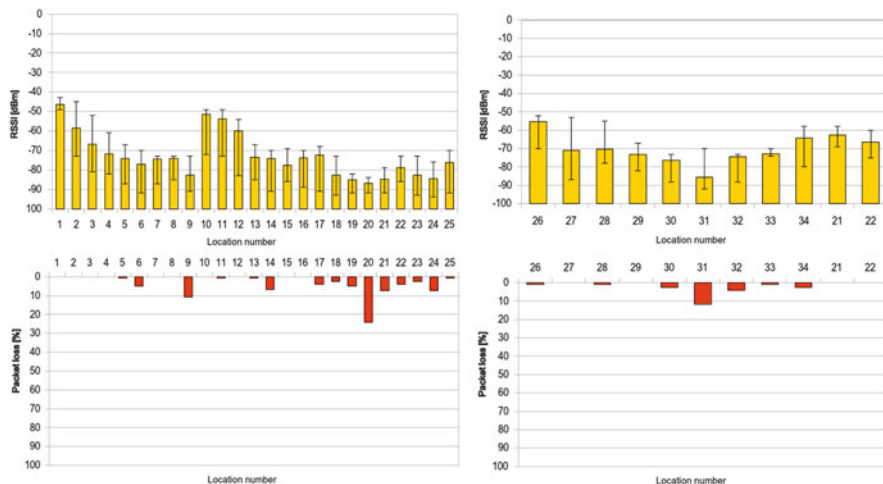
Fig. 13.18 RSSI and packet loss in the center of Logatec measured between light poles and the coordinator's location (*top* and *bottom left*) and measured between light poles and location number 25 (*top* and *bottom right*)

Fig. 13.19 Device types deployed in Logatec industrial zone (*left*) and city center (*right*)

enough to install only a few UHF band TV receivers in the two clusters of the testbed.

Based on the observations from above, the light pole allocations depicted in Fig. 13.19 have been created for the industrial zone and for the city center. On both figures the striped, white and shaded markers with a dot indicate the locations of the 868 MHz SRD band, 2.4 GHz ISM band and the UHF band sensing nodes,

respectively, and the white markers without a dot show locations reserved for future expansion of the testbed. The locations of the two coordinators are not shown explicitly, although their locations can be easily found from Figs. 13.15 and 13.17 and they coincide with the locations of the UHF band sensing nodes.

As it can be seen from Fig. 13.19, three VESNA nodes with SNE-ISMTV-UHF receivers are planned for the industrial zone, and four for the city center. The locations of the TV receivers are evenly distributed along the area of the testbed. For the rest of the locations, in the industrial zone the pattern of one sensor for 868 MHz SRD band and two sensors for 2.4 GHz ISM band is repeated, while in the center the 868 MHz and 2.4 GHz ISM band sensors are assigned to light poles in an alternating manner. This way 14 sensors for 2.4 GHz ISM band were installed in the industrial zone and 15 in the city center, whereas for 868 MHz SRD band sensing 9 and 7 sensors were installed in the industrial zone and in the city center, respectively. The reserved locations are marked on the maps because the public light poles are powered in groups, so the reserved locations will have to be powered in any case. Knowing that power is already available at some locations is useful for the future expansion of the testbed.

13.3.5 *LOG-a-TEC Testbed Access, Control and Reconfiguration*

Remote access to and configuration of nodes in a testbed is supported through the wired or wireless management network. In case of the VESNA-based testbed the management network among nodes is based on ZigBee, as implemented in the ZigBit port to Atmel modules. It forms a wireless multihop network, which means the gateway is able to communicate also with the nodes out of its direct radio range.

For the purpose of communication between sensor network and the remote server (infrastructure side) we developed a new protocol (see Fig. 13.20), which was inspired by the HTTP protocol and is simple enough for fast implementation on VESNA nodes. The protocol defines two types of requests, GET and POST, which are understood by every VESNA node. The GET is used for "safe" requests which do not change the state of the system and POST for "unsafe" requests which change the state of the system. The response from a node following a GET request is considered to be in binary format and handled accordingly, although general responses are in text format and only the spectrum sensing data is in binary format. The ending of each response is indicated with the sequence OK\r\n.

The protocol includes simple but efficient error handling mechanism. There are two types of errors defined. The first is JUNK-INPUT, which is the more common situation when the resource name is mistyped and the parser on the node does not recognize it. The second type of error is CORRUPTED-DATA, used when cyclic redundancy check (CRC) check did not succeed thus we can conclude that the error happened somewhere on the line between the infrastructure and the gateway. The last situation will occur with very low probability.

```
GET:
    GET resource?arg1=val1&arg2=val2&...&argN=varN\r\n

        resource:    abstract resource identifier
        examples:    - firmare/version
                     - sensors/temperature
        arg1:        parameter 1 name
        val1:        value of parameter 1
        ãÅȩ
        argN:        parameter N name
        valN:        value of parameter N

POST:
    POST resource?arguments\r\n
    Length=len\r\n
    <data, having len bytes length>\r\n
    crc=crc_value\r\n

        resource:    abstract resource, e.g. firmware
        arguments:   arguments given to the handler of POST
        example:     2.34/firmware
        len:         length of the data that will be written to
                     the specific resource
        data:        possibly binary data, to be transmitted
        crc_value:   value of CRC calculated on all the previous
                     content except the line starting with crc=;
                     value represented as an unsigned decimal
                     number

Responses from the coordinator have the general form:

    <response to a specific request>\r\n
    OK\r\n
```

Fig. 13.20 Resource access protocol

The protocol is designed as a client-server protocol. In our case the server side is on sensor nodes and the client side is on the remote server. Before we can access the resources the gateway has to establish a connection with the infrastructure side. This is done by establishing a secure SSL encrypted socket with the remote server. The gateway has an Ethernet module embedded on the expansion board which is used to connect the gateway to the internet. The Ethernet module is configured to get the IP address from DHCP server and then automatically tries to set up an encrypted SSL socket with one of the SSL servers listening on a specific port located on the remote server. Once the connection has been established one could access any resource (sensor, radio module, etc.) or procedure on any of the nodes. Procedures pre-prepared on the nodes include remote reprogramming, start spectrum sensing, collect spectrum data, configure nodes as transmitters, configure frequency band, etc.

13.3.6   Summary

This section presents the testbed and management network planning performed for the LOG-a-TEC testbed, taking into account high-level goals of the testbed and the constraints and performance requirements of the management network. The methodology for performance evaluation of the network used in the planning of

the testbed is presented, and results from actual field measurements are provided and analysed, justifying the selection of device locations and their distribution in the LOG-a-TEC testbed as per sensing functionality. Section concludes with the description of the protocol used for remote access, control and configuration of the LOG-a-TEC testbed making use of the underlying management network based on Atmel's ZigBit network.

13.4 Signal Generation Use Case

The LOG-a-TEC testbed presented in previous sections can be used for different experiments in communications, facilitating the transition of technology from fully controlled computer simulation environments to actual prototyping. As explained in Sect. 13.2, the main role of VESNA nodes equipped with the SNE-ISMTV expansion boards is spectrum sensing, and as deployed in the LOG-a-TEC testbed they can perform distributed spectrum sensing. However, investigation of primary-secondary scenarios requires also controlled emulation of primary users, especially wireless microphones. To remove the need to manually introduce wireless microphones into the testbed during the experimentation, VESNA nodes can also support remotely controlled emulation of wireless microphone transmissions. In particular, this has been enabled on VESNA nodes equipped with the SNE-ISMTV-TI868 transceiver, that is capable of transmitting a narrow-band signal at frequencies between 780 and 862 MHz. This frequency span includes the upper UHF channels that are used by licensed wireless microphone users. In LOG-a-TEC, one or more of these nodes can be remotely triggered to act as wireless microphones during the experiment or demonstration.

To create as realistic environment as possible in the testbed it was decided to use the IEEE wireless microphone emulation profiles [18]. This method approximates a radio signal transmitted by a legacy analogue studio wireless microphone (PMSE) using a frequency modulated continuous sine wave $s(t)$.

$$s(t) = A \cos \left(2\pi f_c t + \frac{f_{dev}}{f_m} \cos \left(2\pi f_m t \right) \right) \tag{13.1}$$

In Eq. (13.1), A is the carrier amplitude, f_c is the carrier frequency (between 780 and 862 MHz), f_{dev} is the frequency deviation and f_m is the frequency of the continuous sine wave base band signal. Three signal profiles are defined in literature as summarised in Table 13.4, i.e. silent, soft speaker and loud speaker. Each profile corresponds to one typical operating condition of the wireless microphone and defines the values for f_{dev} and f_m.

The SNE-ISMTV-868 transceiver is not directly capable of transmitting a modulated analogue signal as required by this method. However it contains a digital frequency-shift keying modulator block with 4 symbols (4FSK) that can be used to transmit a continuous signal. In FSK each symbol is represented by a sine with a

Table 13.4 FM parameters for wireless microphone simulation profiles

| Operating mode	f_m (kHz)	f_{dev} (kHz)
Silent	32.0	5.0
Soft	3.9	15.0
Loud	13.4	32.6
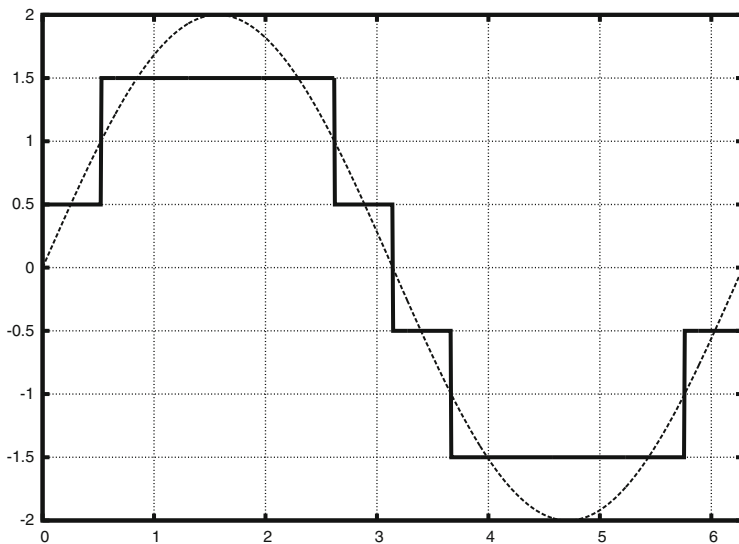
Fig. 13.21 Approximating a sine wave with frequency-shift keying transmission. *Dotted line* depicts an ideal analogue signal while the *bold line* shows an approximation achievable with a 4FSK modulator

frequency that is shifted from the carrier frequency by a discrete amount, determined by the symbol value. When fed with an appropriate stream of symbols at a high enough rate, such a modulator can be used to approximate an analogue waveform, as depicted in Fig. 13.21.

Using the 4FSK modulator an approximation was hence implemented of the wireless microphone emulation profiles. The block diagram of the implementation is depicted in Fig. 13.22. The transceiver was configured for 200 ksymbols/s, which was the maximum symbol rate that was achievable on the VESNA core CPU running at 64 MHz. This is was equivalent to a transmission of a sampled analogue signal with 2-bit quantization and $f_s = 200$ kHz sampling rate. The continuous symbol stream was provided by the VESNA core CPU to the transceiver through a synchronous serial bus.

To generate the appropriate symbol stream for the modulator hardware block a direct digital synthesis (DDS) algorithm was implemented in software on the CPU. A DDS consists of a phase accumulator, a tuning word *TW* register and a table of length *L* that maps between phase value in the desired signal amplitude. For each signal amplitude value emitted by the DDS algorithm, the value of the tuning word
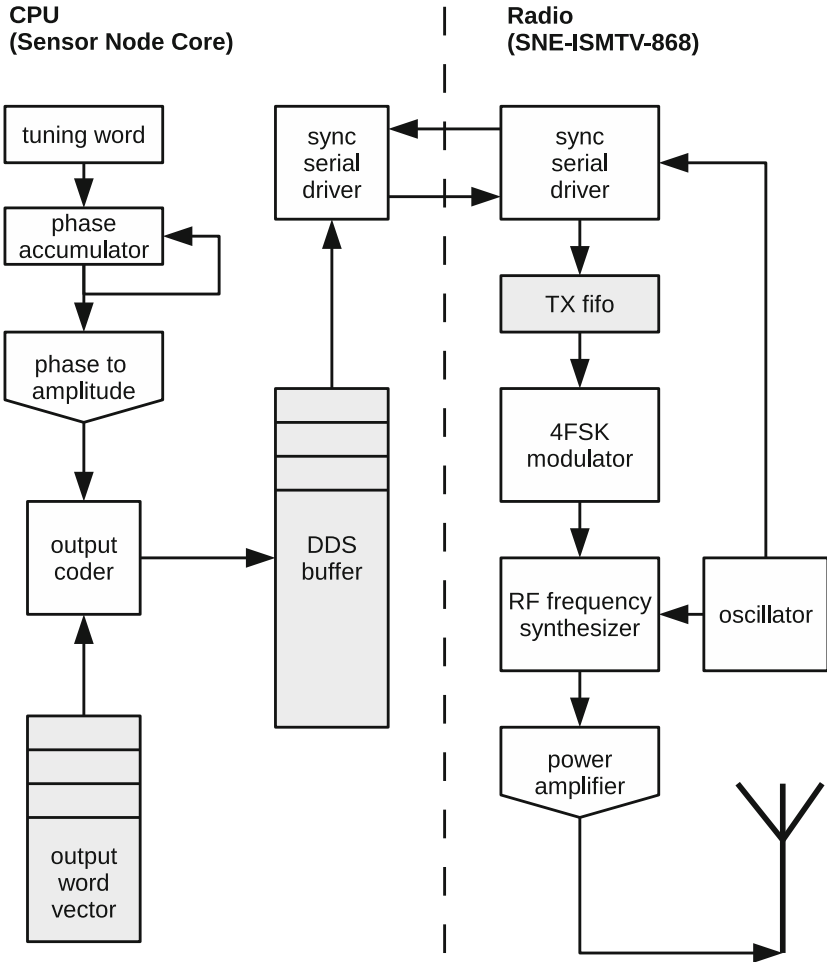
Fig. 13.22 Block diagram of direct signal synthesis on VESNA with SNE-ISMTV expansion

is added to the phase accumulator. Hence the frequency of the baseband signal f_m can be adjusted by software simply by setting an appropriate tuning word value, according to Eq. (13.2).

$$TW = L \frac{f_m}{f_s} \tag{13.2}$$

In our case an output coder was also required to map the signal amplitude to the appropriate 4FSK symbol. Best mappings between signal amplitude and 4FSK symbol were determined experimentally, as they were not provided by the manufacturer of the transceiver.
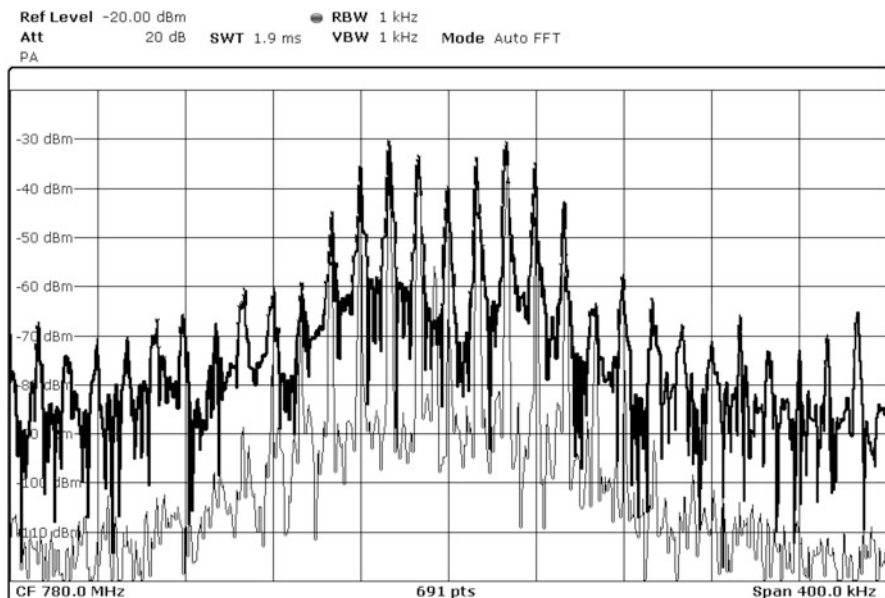
Fig. 13.23 Spectrum of a "loud speaker" wireless microphone simulation signal produced by SNE-ISMTV (upper, thick trace) compared to the same signal produced by a USRP device (lower, thin trace)

To set the frequency deviation f_{dev} according to Table 13.4, the hardware FSK modulator deviation setting was adjusted by writing to the appropriate transceiver hardware register. By adjusting these two settings (FSK deviation and tuning word) we were able to closely match the FM parameters required for all three wireless microphone simulation profiles. An example of the spectrum for the VESNA-emulated wireless microphone signal corresponding to a loud speaker compared to the same signal produced by a USRP device is depicted in Fig. 13.23.[2]

The three supported wireless microphone emulation profiles are available in all VESNA nodes with the SNE-ISMTV-TI868 transceiver in the LOG-a-TEC testbed and were seamlessly integrated with the signal generation API, which is exposed by sensor nodes. To cover the entire UHF frequency band two configurations were required for each emulation profile due to the limitation on the number of channels per hardware transceiver configuration (256). Hence when queried for signal generation profiles (see Fig. 13.24), these nodes report 6 wireless microphone signal generation configurations.

---

[2]The program source code for the wireless microphone emulation is available at https://github.com/avian2/vesna-audio-synthesis

```
GET nodes?8/generator/deviceConfigList
dev #0, CC1100, 9 configs:
cfg #0: CC1100, FM noise, 200 kHz deviation:
    base: 779999908 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 256,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #1: CC1100, FM noise, 200 kHz deviation:
    base: 829999924 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 160,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #2: CC1100, wireless mic, silent:
    base: 779999908 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 256,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #3: CC1100, wireless mic, silent:
    base: 829999924 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 160,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #4: CC1100, wireless mic, soft speaker:
    base: 779999908 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 256,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #5: CC1100, wireless mic, soft speaker:
    base: 829999924 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 160,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #6: CC1100, wireless mic, loud speaker:
    base: 779999908 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 256,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #7: CC1100, wireless mic, loud speaker:
    base: 829999924 Hz, spacing: 199814 Hz, bw: 197754 Hz, channels: 160,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
cfg #8: CC1100, 868 MHz SRD band, FM noise, 50 kHz deviation:
    base: 863999695 Hz, spacing: 49953 Hz, bw: 49438 Hz, channels: 140,
    min power: -30 dBm, max power: 12 dBm, time: 5 ms
```

Fig. 13.24 The profiles currently supported by the VESNA SNE-ISMTV-TI868 in LOG-a-TEC

13.4.1 Summary

SNE-ISMTV-UHF is a low cost custom designed receiver only and therefore unable to transmit in VHF and UHF bands. In order to support the emulation of wireless microphone for LOG-a-TEC, the SNE-ISMTV-TI868 extension board was used to enable generating wireless microphone profiles that are compliant with the IEEE requirements. This functionality can be used in a coexistence study of incumbent and cognitive radio systems in the TV white spaces, as for example in the CREW-TV project concerned with the investigation of the benefits of combining a TVWS database with a distributed sensing network.

13.5 An Interference Mitigation Use Case

The use case presented in Sect. 13.4 demonstrates how the LOG-a-TEC testbed, or even individual VESNA nodes, can be used to support the investigation of coexistence of primary and secondary users in TVWS, while the use case in this section demonstrates the use of the testbed for investigation of approaches to coexistence of secondary devices in shared radio spectrum. In such use case the experimenter is faced with real-world problems and challenges with the implementation of the investigated approach or algorithm that have to be appropriately addressed. In the following we show an example of a game theoretic approach to cognitive radio with power allocation and interference mitigation game with more detailed insights available in [19].

Game theoretic approaches to the problem of resource management in wireless communications have been extensively studied over the last decade as they provide a more flexible and dynamic alternative to co-existence and resource sharing. One of the simplest forms of co-existence is coordinating the interference among several transmitters by limiting the transmit power. For instance, in [20], the authors propose a theoretical framework for a game-theoretic power control scheme for wireless and ad-hoc networks. In [21], the authors reformulate the problem by also considering energy efficiency alongside with interference and propose the ProActive Power Update (PAPU) algorithm. Simulation results showed that PAPU can utilize energy more efficiently by sacrificing a small amount of network utility compared to the Asynchronous Distributed Pricing (ADP) protocol proposed in [20]. An example use case enabled by the LOG-a-TEC testbed is to evaluate the PAPU algorithm in a realistic outdoor scenario.

13.5.1   Summary of the PAPU Algorithm

The power allocation game proposed in [21] is formulated as follows. Given a wireless network of N transmit-receive pairs (Tx_i-Rx_i), where a "pair" is referred to as a "player", the objective is to find stable points of power allocation for each player such that the players' global utility is maximum while the cumulated power levels are kept to a minimum.

More formally, given a set of N players, $N = \{1, 2, \ldots, N\}$, and their corresponding power allocation profile $P = \{p_1, p_2, \ldots, p_N\}$, the utility function of each player is given by:

$$u_i = log(1 + \frac{h_{ii} p_i}{n_0 + \frac{1}{B} \sum_{j \neq i} h_{ji} p_j}) \tag{13.3}$$

where p_i, p_j are the transmit powers of players i and j, h_{ii} is the direct gain, h_{ji} is the channel gain between transmitter j and receiver i, B is the total channel bandwidth and n_0 is the noise power. For simplicity and in accordance to [21], we consider B = 1. Now, the objective is to maximize the global utility function (13.4a), while minimizing the globally allocated power (13.4b) where $p_i \in [0, P_i^{max}]$.

$$max \sum_i u_i \tag{13.4a}$$

$$min \sum_i p_i \tag{13.4b}$$

In the evaluations performed in a simulation environment [21], the values for the direct gain, the cross gains, the transmitted power and noise were chosen

conveniently to provide the proof of concept. Communication and reconfiguration delay were not taken into account.

The validation of theoretical models in real-world set-ups poses several constraints that are most often testbed specific. We employed the following methodology for investigating the feasibility of experimenting with interference mitigation based on the power allocation game in LOG-a-TEC:

- Identification of the experimental set-up and the constraints.
- Adaptation of the theoretical framework for the use in a testbed rather than in a simulation scenario.
- Empirical determination of the values of parameters such as channel gain.
- Implementation and experimental evaluation.

13.5.2 Experimental Set-Up

After identifying and understanding the theoretical framework we desire to validate, the next step is to understand the constraints imposed by the testbed we plan to use. These constraints will require an adaptation of the framework so that it can be validated in a realistic set-up. For the case of LOG-a-TEC, the following constraints can be identified:

- **Topological** – The topology of the testbed is determined by the alignment of the public light poles on which the sensor nodes are mounted. The theoretical framework behind PAPU assumes that the cross gains are significantly smaller than the direct gains. If this constraint is not satisfied, the game might not converge. This constraint and the topology of the testbed narrows down the choice for the location of the players.
- **Transmission capability** – The testbed nodes are able to transmit on one channel at a time, therefore limiting the type of games that can be supported to single channel ones.
- **Power levels** – The nodes' CC2500 transceivers support discrete power levels. Unlike in the theoretical framework, where continuous power levels are considered, the empirical game has to be adjusted to one of the associated power levels specified by the radio chip. This clearly affects players' strategies as well as the resulting stable points of power allocation for players.
- **Sensing** – The nodes in the testbed use energy detection for spectrum sensing. This simple method cannot distinguish between different types of detected signals. As a result, the accuracy of the measurements is lower, therefore transmit power level adaptations of the players may be misguided by errors.
- **Delay** – The delay for setting up a transmission or a sensing vary depending on the nodes, since the management network through which this setup is conducted is wireless. Typical values for this setting are from 1 to 3 s. The delay affects the speed of the game, thus the time required to converge.
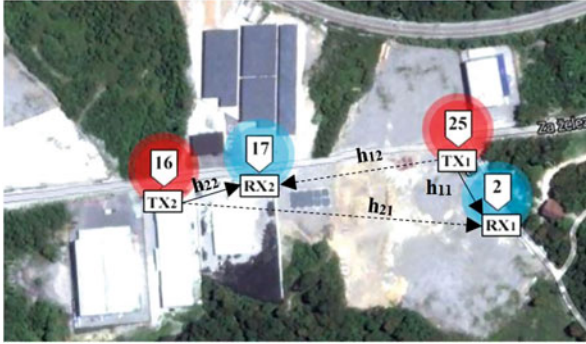
Fig. 13.25 Inter-network interference between wireless systems where h_{ij} is the channel gain between transmitter i and receiver j. Pentagonal markers point to sensor node locations. Numbers in markers are location identifiers

- **Synchronization** – The nodes in the testbed do not use a clock synchronization protocol, therefore the lack of synchronization has to be taken into account when designing and implementing the game.
- **Calibration** – The low cost nodes are not calibrated, therefore a setting of 0 dBm transmission power might result in a slight shift of the level of the actual transmitted power. This affects the final strategies of the players.

Considering the theoretical framework behind the PAPU algorithm and the constraints imposed by the LOG-a-TEC testbed we defined the interference-aware power control game between two players operating at 2.4 GHz, in the industrial zone cluster. The discrete transmit power levels for the VESNA nodes in LOG-a-TEC can be set to [0, −2, −4, −6, −8, −10, −12, −14, −16, −18, −20, −22, −24, −26, −28, −30, −55] dBm. The two players are transmit-receive pairs coexisting in the same area, as depicted in Fig. 13.25. Player 1 is formed by the Tx-Rx pair Node25-Node2 whereas player 2 is formed by the Tx-Rx pair Node16-Node17. The distances between the nodes are: $d(TX_1, RX_1) \approx 50$ m, $d(TX_2, RX_2) \approx 65$ m, $d(TX_2, RX_1) \approx 230$ m, $d(TX_1, RX_2) \approx 150$ m.

13.5.3   The Adaptation of the Theoretical Framework

The game adopted in this use case is based on players (users) adapting their transmit power level when detecting a change in other players' power level. The new transmit power level, to which a node adapts, is called the best response. The best response of any of the players involved in the game is given by [21]:

$$b_i(p_{-i}) = \frac{1}{c_i} - \frac{\sum_{j \neq i} h_{ji} p_j + n_0}{h_{ii}} = \frac{1}{c_i} - \frac{I + n_0}{h_{ii}} \tag{13.5}$$

where $b_i(p_{-i})$ represents the best response of player i given the current state of the game (the power profile for all other players is denoted by p_{-i}), c_i represents player i's energy cost, h_{ij} are the channel gains, p_j is the transmitted power for all the other players and n_0 is the noise. In typical controlled evaluation setups, such as simulators, the values for h_{ji} and p_j are chosen conveniently to guarantee convergence of the game. In a real-world testbed, however, these parameters have to be acquired from the operating environment and may influence the outcome of the game. As a result, the following adapted expressions were used in the experimental evaluation:

$$b_i(p_{-i}) = \frac{1}{c_i} - \frac{\sum_{j \neq i} h_{ji} p_j + n_0}{h'_{ii}} = \frac{1}{c_i} - \frac{Pr_{measured}i - P_{useful}}{h'_{ii}} \qquad (13.6)$$

$$b_i(p_{-i}) = \frac{1}{c_i} - \frac{\sum_{j \neq i} h_{ji} p_j + n_0}{h'_{ii}} = \frac{1}{c_i} - \frac{Pr_{measured|p_i=0}}{h'_{ii}} \qquad (13.7)$$

where h'_{ii} stands for the measured or the measurement-based estimation of the direct gain, $Pr_{measured}i$ stands for the received power measured by player i when player i is also transmitting ($p_i \neq 0$), P_{useful} stands for the estimated useful power received by Rx_i when Tx_i is transmitting, and $Pr_{measured|p_i=0}$ stands for the received power measured by player i when player i's transmitter is silent ($p_i = 0$).

For the entire system to be stable, the PAPU algorithm must converge to a Nash equilibrium. The convergence issue studied in [21] gives the following condition for the convergence and stability of PAPU:

$$\left| \frac{h_{ji}}{h_{ii}} \right| < \frac{1}{N}, i = 1, \ldots, N \qquad (13.8)$$

Eq. (13.8) is a decisive factor when choosing the topology of nodes in the testbed on which the power allocation game can be implemented. If Eq. (13.8) is not fulfilled for all players, there will be no strategy profile that will satisfy the players.

In the theoretical case, considering Eq. (13.5), an equilibrium is reached in the game when $p_i(t - 1) = p_i(t)$ for all players at once. However, in practice, where p_i can take only discrete values, a more robust stopping criterion is needed in order to determine the equilibrium. In this study, the criterion is given by the following condition: $|b_{i-k}(p_{-i}) - b_i(p_{-i})| < P_{th}$, where $k = \overline{0, 4}$ and P_{th} stands for the threshold power used to compensate for the environment dynamics.

Based on the PAPU algorithm, we defined the following power control protocol:

- Step 1: Each player i initializes its power p_i^0, p_{-i}^0.
- Step 2: At time t if player i updates its power, player i will alert the neighbors that a power change has been made.
- Step 3: If player i detects a change in other players' power, it updates its power according to Eq. (13.7) and alerts the neighbors of the change.
- Step 4: Player i checks if the Nash equilibrium condition is satisfied
- Step 5: If the Nash equilibrium condition is satisfied, the game is stopped.

The code implementing this protocol is too complex to include in this chapter, however it is publicly available in our repository on GitHub.[3]

13.5.4 Empirical Parameter Determination

The channel estimation for the topology presented in Fig. 13.25 was performed starting from empirical data. The gain between pairs of nodes has been measured and three strategies were investigated for deriving their values for the purpose of the PAPU algorithm, i.e. average gain, instantaneous gain and estimated gain using Kalman filter.

The channel gain was measured as follows. First, the receiver Rx_j measured the received power P_{noise} knowing that the transmitter Tx_i was silent using the code presented in Fig. 13.26 lines 1–9 and 15–23. Then, the receiver measured the received power P_{Rx_j} knowing that the transmitter was transmitting with a given power P_{Tx_i} using the code in Fig. 13.26 lines 1–9, 11–19 and 25–27. The gain (lines 29–31 in Fig. 13.26) was then computed according to:

$$h_{ij} = \frac{P_{Rx_j} - P_{noise}}{P_{Tx_i}} \tag{13.9}$$

Long term channel gain measurements were performed for 19 consecutive days, between August 5th and August 23rd 2013, two times a day, in the afternoon (12:30–13:30) and late evening (21:00–22:00). For channel gain measurements, a signal at 2420 MHz with a bandwidth of 400 kHz was used at the transmitter side. The measured direct channel gains for player 1 are represented in a chronological order in the plot depicted in Fig. 13.27. The vertical white and grey shadings represent the afternoon period with white and the late evening period with gray, respectively.

As shown in Fig. 13.27, the average gain for the pair of nodes representing Player 1 is −74.8 dB with a standard deviation of 6.952 dB and a dynamic range of 18.9 dB for the 19-day period considered. Even in short term measurement periods, such as an afternoon or an evening session, gain variations of over 5 dB can be found. Similar observations can be made for other pairs of nodes. The gain is highly dynamic and thus using the average as a means of approximation does not reflect the actual state of the channel. Therefore the best-response strategy of the power allocation game will yield unrealistic values. This method would be suitable if the gain variation were not considerable (e.g. <3 dB), which happens only occasionally.

Alternative to average gain as input to the best-response strategy is the instantaneous channel gain. This approach, however, can prove expensive in terms of delay

---

[3]https://github.com/sensorlab/logatec-games/blob/master/Power_Allocation_Game/powerAllocation.py

```
1.  # select the pair of nodes
2.  tx_node = Node(coordinator_id, tx_node_id)
3.  rx_node = Node(coordinator_id, rx_node_id)
4.
5.  # settings
6.  p_tx_dBm = 0
7.  sensing_duration = 2
8.  time_after = 1.5
9.  attempts = 0
10.
11. # these lines are needed just for signal measurement
12. tx_node.setGeneratorConfiguration(measuring_freq, p_tx_dBm)
13. tx_node.generatorStart(time.time(), transmitting_duration)
14.
15. #configure rx_node
16. rx_node.setSenseConfiguration(measuring_freq,
17.     measuring_freq, 400e3)
18. #start sensing
19. rx_node.senseStart(time.time()+time_after, sensing_duration, 5)
20.
21. #these lines are needed just for noise computation
22. noise_power = GainCalculations.getAverageDataMeasurementsFromFile(
23.     coordinator_id ,rx_node.node_id)[1][0]
24.
25. #these lines are needed just for noise computation
26. received_power = GainCalculations.getAverageDataMeasurementsFromFile(
27.     coordinator_id,rx_node.node_id)[1][0]
28.
29. #these lines are needed just for gain computation
30. gain = (received_power - noise_power) /
31.     (math.pow(10.00, p_tx_dBm/10.00) * 0.001)
```

Fig. 13.26 Example code required to measure the noise between a pair of nodes (Note: This is example code, the actual working code is available at [22])

(for the particular case of the LOG-a-TEC testbed one such measurement takes 5–6 s) and power consumption, and it is prone to measurement errors.

The third approach considered in this use case for deriving the channel gains is to use a simple channel estimation technique such as the Kalman filter. This method consists of two steps: the first step is estimating the next measured gain and the second step represents the correction of the estimated gain by performing a new gain measurement. The resulting predicted gain is then computed. The values of the Kalman predictor for long term measurements are depicted in Fig. 13.27. For a short history, it approximates well the channel gain (see first 3–5 measurements) but it gets worse as the history gets longer.

By analyzing the long-term measurements, three important conclusions reflecting the implementation of the game can be drawn. First, the channel gain Kalman predictor approach is more suitable for the power allocation game because it compensates for the measurement errors to which the instantaneous gain is prone, while also relying on measurement history. Second, the Kalman filter based estimations of the gain are more accurate than the average. This is supported by the smaller values of the mean square error between the Kalman estimations and the instantaneous value (K−inst) than the values of the mean square error between average and instantaneous values (avg−inst) as listed in Table 13.5. Third, using a shorter history for the Kalman predictor yields better results while also having some practical advantages as far as implementation is concerned. At the beginning of each experiment, the Kalman predictor is used with a history of 9 most recent measurements.
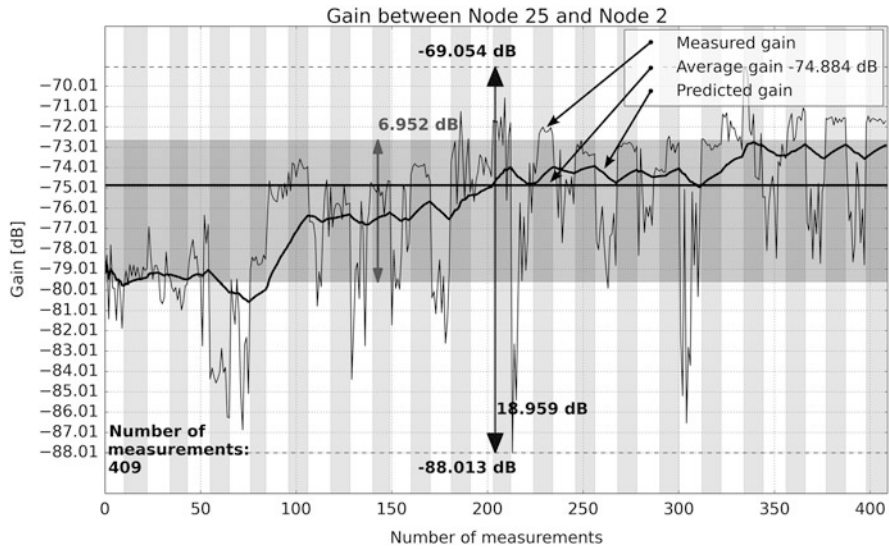
Fig. 13.27 Long term measurements of direct channel gains for player 1 (from August 5th to August 23rd 2013 between 12:30–13:30 – *white vertical shading* and 21:00–22:00 *gray vertical shading*)

Table 13.5 Mean squared error for the average and predicted gain with respect to instantaneous gain

MSE	Average gain (avg–inst)	Predicted gain (K–inst)
h_{11}	4.65×10^{-16}	2.14×10^{-16}
h_{12}	7.9×10^{-17}	1.59×10^{-17}
h_{21}	2.8×10^{-19}	2.55×10^{-19}
h_{22}	1.26×10^{-15}	8.11×10^{-16}

13.5.5   Experimental Results

The setup presented in previous subsections was used for experimental performance evaluation of the PAPU algorithm for power allocation and interference mitigation. In particular, experimental results were obtained to show (i) the effect of energy cost on the best response of the players, (ii) the effect of the channel gains on the best response of the players, and (iii) the influence of real-world operating conditions on reaching a Nash equilibrium.

The best response formula used for the empirical evaluation of the game is given in Eq. (13.5). The first term of the formula is the cost c_i whose role is to penalize the players for transmitting with high power. The feasible values of the cost also have to take into account the power levels supported by the testbed. In the case of LOG-a-TEC, the best response has to respect the lover boundary of -55 dBm and the upper boundary of 0 dBm. This constraint excludes a large set of low values of the cost. Furthermore, the experiments have shown that a very high cost value prevents the game to converge to an equilibrium making the communication system unstable.
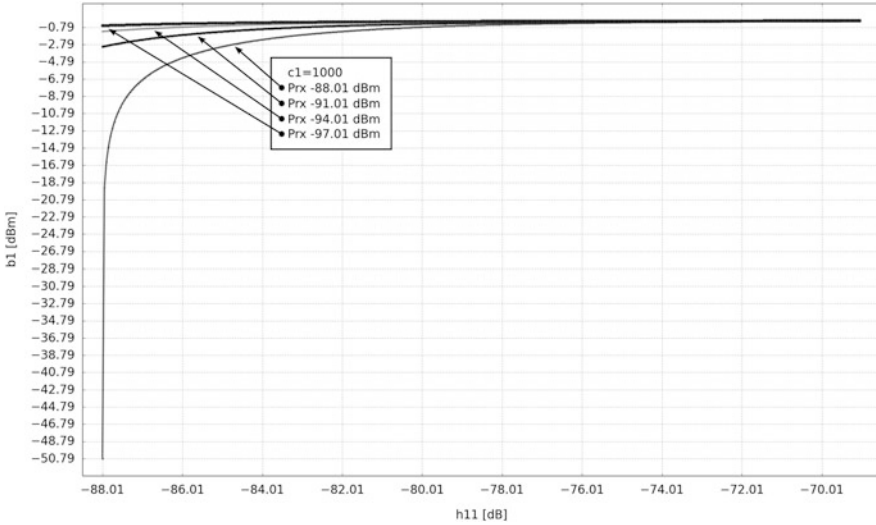
Fig. 13.28 The influence of the channel gains on the best response for Player 1 $b_1(p_{-1})$ versus h_{11} and $I + n_0$

This observation has led to putting a higher boundary on the cost value. In the worst case scenario observed during the experiments, when the $I + n_0$ is -84 dBm for the first player and -72 dBm for the second player, the feasible values for c_i lie in the [1000; 4000] interval.

The gain is another of the parameters of the best response formula from Eq. (13.5) and the logarithmic representation of the dependency of $b_i(p_{-i})$ to h_{ii} for Player 1 is depicted in Fig. 13.28. For a fixed cost of $c_i = 1000$ for both players and the worst case scenario where the values of $I + n_0$ are -88 dBm for player 1, it can be seen that the best response has high variations for small gains and negligible variations for high ones.

The existence of a Nash equilibrium for the PAPU algorithm has been proven in [21]. The simulated value of the Nash equilibrium for PAPU considering $c_i = 1000$ and average value of the gain is $(-0.24, -1.25)$ as can be seen in Fig. 13.29. By introducing more realistic gains based on the Kalman estimator, the simulated Nash equilibrium slightly varies around $(-0.14, -1.1)$, as depicted with triangles in Fig. 13.29. In this case the influence of the predicted gains to the player's final strategies is evident.

By running the game on the real−world testbed, the Nash equilibria are more spread as shown with squares in Fig. 13.29, mostly due to interference and noise. Still, the variation of the best response is relatively small with 0.3 dBm for Player 1 and 0.5 dBm for Player 2. In the first two cases where simulation was used, the Nash equilibrium results in a $(0, -2)$ value. For the case of LOG-a-TEC, in the experimental setup, the multitude of Nash equilibria obtained in different runs, after
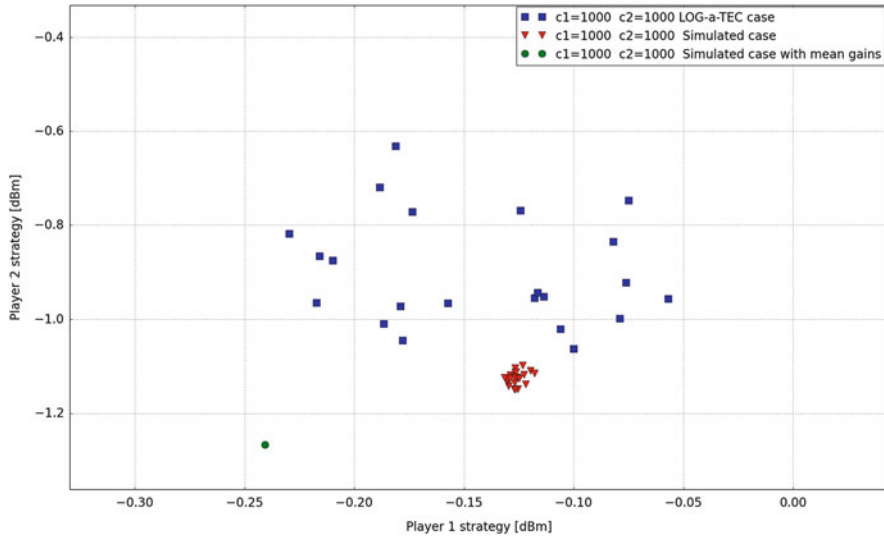
Fig. 13.29 Experimental Nash equilibrium compared to simulated one for two players
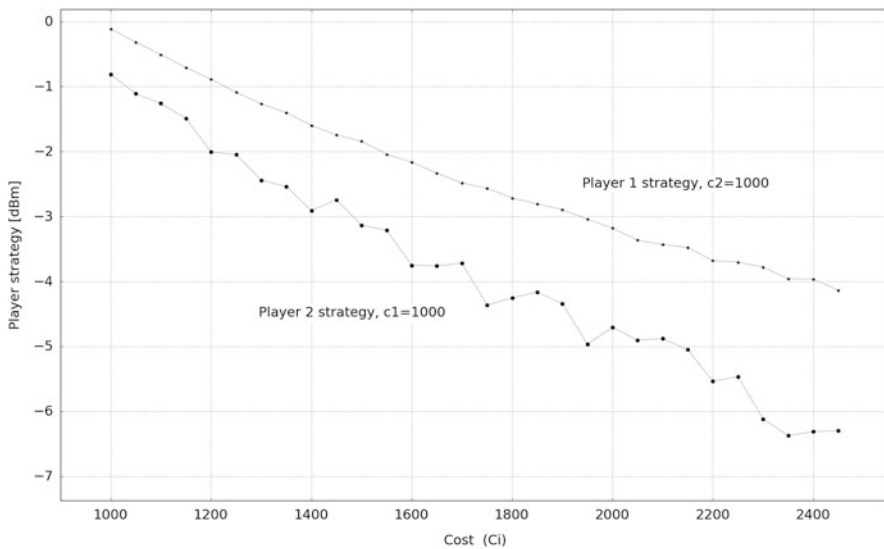

Fig. 13.30 Players' best response as function of players' costs

being rounded to the discrete values of the transmission power supported in LOG-a-TEC, are (0,0) and (0, −2).

Figure 13.30 depicts player strategies for different cost schemes. It can be seen that with the increase of the cost c_i, there is a decrease of the allocated powers. All the experiments were performed as following: the channel gains h_{ii} were computed

at the beginning of each experiment by predicting the gain, interference and noise power levels. $I + n_0$, were measured and updated at each iteration, $k = \overline{0,4}$ and $P_{th} = 0.8$ dBm. For Fig. 13.30, all the experiments were repeated 20 times and the Nash equilibrium are given as averaged values.

13.5.6 Summary

This section demonstrated the role of a testbed such as LOG-a-TEC in validating techniques proposed for cognitive radios with more detailed insights available in [19]. We selected the ProActive Power Update algorithm based on game theoretic framework, and showed how it can be adapted for validation on LOG-a-TEC. Besides validating the proposed framework and algorithms, the effort enabled studying the effects of the empirical parameter estimation on the best response and players' strategies which represent the Nash Equilibria. The results showed that for a certain cost range, the system can reach the Nash equilibria. The equilibria and the convergence time are strongly influenced by the cost but also by the channel gains, requiring special care in selecting the topology of nodes within the testbed.

These results may prove useful in developing new protocols in decentralized CR networks and the current implementation is openly available in our repository on GitHub.[4]

13.6 Summary of the Chapter

In this chapter we described in our view and based on our experience the most important of the aspects related to the design and deployment of the LOG-a-TEC embedded, outdoor cognitive radio testbed. We started the chapter by discussing radio front-end for embedded testbeds and described the SNE-ISMTV expansion board custom designed for the VESNA sensor node for supporting cognitive radio experimentation with low cost low-power devices. The expansion board can be equipped with off-the-shelf reconfigurable transceivers for experimentation in license free SRD and ISM bands at 868 and 2400 MHz, respectively, or with more sophisticated custom designed hardware operating in VHF and UHF TV bands. The chapter continues with practical experiences for designing the testbed infrastructure, like the management network and our considerations in the choice of network protocols employed in the LOG-a-TEC testbed.

After describing all these aspects of LOG-a-TEC, we provide two use cases that can be supported by the testbed. The first one consists of generating a wireless microphone profile for realizing experiments in UHF bands where the corresponding SNE-ISMTV-UHF is only capable of receiving. The wireless microphone

---

[4]https://github.com/sensorlab/logatec-games

profiles are generated using the SNE-ISMTV-TI868 extension and can be used for instance to perform coexistence experiments where the spectrum sensing data is fed into a hybrid geolocation database. The second use case consists of the validation of an interference mitigation strategy using a game theoretic framework. The discussion includes the adaptations required by the theoretical framework as well as discussion with respect to the constraints posed by the testbed operating in real-world environment. An evaluation of the approach and discussion related to the trade-offs that were observed is also provided.

In summary, the chapter gives a quick insight into the potential for experimentally driven research held by embedded testbeds consisting of low cost low-power devices. At the same time it points out some important choices one needs to make in the process of the design, which have significant impact on subsequent usability and functionality of the testbed. These especially include the issues related to topology design, power supply, network management, remote access and reconfigurability of devices.

References

1. TmoteSky ultra low power IEEE 802.15.4 compliant wireless sensor module. http://www.eecs.harvard.edu/~konrad/projects/shimmer/references/tmote-sky-datasheet.pdf. Jan 2014
2. VESNA sensor node. http://sensorlab.ijs.si/hardware.html. Jan 2014
3. Dunkels, A., Gronvall, B., Voigt, T.: Contiki-a lightweight and flexible operating system for tiny networked sensors. In: 29th Annual IEEE International Conference on Local Computer Networks, 2004, Tampa, Florida, USA, pp. 455–462. IEEE (2004)
4. Levis, P., Madden, S., Polastre, J., Szewczyk, R., Whitehouse, K., Woo, A., Gay, D., Hill, J., Welsh, M., Brewer, E., et al.: TinyOS: an operating system for sensor networks. In: Ambient Intelligence, pp. 115–148. Springer, Berlin/New York (2005)
5. CREW – Cognitive Radio Experimentation World. http://www.crew-project.eu/. Jan 2014
6. Blossom, E.: GNU radio: tools for exploring the radio frequency spectrum. Linux J. **2004**(122), 4 (2004)
7. Sutton, P.D., Lotze, J., Lahlou, H., Fahmy, S.A., Nolan, K.E., Ozgul, B., Rondeau, T.W., Noguera, J., Doyle, L.E.: Iris: an architecture for cognitive radio networking testbeds. IEEE Commun. Mag. **48**(9), 114–122 (2010)
8. Šolc, T.: SNE-ISMTV: VESNA wireless sensor node expansion for cognitive radio experiments. In: Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013), Ilmenau, Germany, pp. 1–2. VDE (2013)
9. Boano, C.A., Brown, J., He, Z., Roedig, U., Voigt, T.: Low-power radio communication in industrial outdoor deployments: the impact of weather conditions and ATEX-compliance. In: Sensor Applications, Experimentation, and Logistics, pp. 159–176. Springer, Berlin/Heidelberg (2010)

10. Lanzisera, S., Mehta, A.M., Pister, K.S.J.: Reducing average power in wireless sensor networks through data rate adaptation. In: IEEE International Conference on Communications (ICC'09), Dresden, Germany, 2009, pp. 1–6. IEEE (2009)
11. Mohorcic, M., Smolnikar, M., Javornik, T.: Wireless sensor network based infrastructure for experimentally driven research. In: Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013), Ilmenau, Germany, pp. 1–5. VDE (2013)
12. LOG-a-TEC Cognitive radio. http://log-a-tec.eu/cr.html. Jan 2014
13. Hrovat, A., Ozimek, I., Vilhar, A., Celcer, T., Saje, I., Javornik, T.: Radio coverage calculations of terrestrial wireless networks using an open-source grass system. WSEAS Trans. Commun. **9**(10), 646–657 (2010)
14. Šolc, T., Padrah, Z.: Network design for the LOG-a-TEC outdoor testbed. In: 2nd International Workshop on Measurement-based Experimental Research, Methodology and Tools, Dublin, Ireland (2013)
15. ZigBit 700/800/900 MHz Wireless Modules datasheet. http://www.atmel.com/Images/doc8227.pdf. Jan 2014
16. Atmel AVR2050: Atmel BitCloud Developer Guide. http://www.atmel.com/images/doc8199.pdf. Jan 2014
17. AVR Low Power 700/800/900 MHz Transceiver for IEEE 802.15.4-2006, IEEE 802.15.4c-2009, ZigBee, 6LoWPAN, and ISM Applications Data Sheet. http://www.atmel.com/images/doc8168.pdf. Jan 2014
18. Clanton, C., Kenkel, M., Tang, Y.: Wireless microphone signal simulation method. IEEE 802.22-07/0124r0 (2007)
19. Anton, C., Toma, A., Cremene, L., Mohorcic, M., Fortuna, C.: Power allocation game for interference mitigation in a real-world experimental testbed. In: 2014 IEEE International Conference on Communications, Sydney, Australia. IEEE (2014)
20. Huang, J., Berry, R.A., Honig, M.L.: Distributed interference compensation for wireless networks. IEEE J. Sel. Areas Commun. **24**(5), 1074–1084 (2006)
21. Fang, G., Dutkiewicz, E., Yu, K., Vesilo, R., Yu, Y.: Distributed inter-network interference coordination for wireless body area networks, IEEE Globecom 2010, Miami, Florida, USA. pp. 1–5 (2010)
22. Sensorlab: Logatec code repository. https://github.com/sensorlab/logatec-games/blob/master/gain-computation/gainCalculations.py. Accessed Feb 2014