

Enabling Sensing-based Opportunistic Spectrum Re-usage with Secondary QoS Support

vorgelegt von
Diplom-Ingenieur
Daniel Willkomm
aus Berlin

von der Fakultät IV — Elektrotechnik und Informatik —
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
— Dr.-Ing. —

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Hans-Ulrich Heiß
Berichter: Prof. Dr.-Ing. Adam Wolisz
Berichter: Prof. Luiz DaSilva, Ph.D.

Tag der wissenschaftlichen Aussprache: 17. Mai 2011

Berlin 2011
D83

Acknowledgments

Throughout the years this Ph.D. thesis has been written I was supported, guided, and accompanied by many people. Without all those people the project “*Dr.-Ing. Willkommen*” would not have been possible.

First of all I would like to thank my advisor Prof. Wolisz for the guidance and (financial) support throughout all those years. Although it was often painful, I admire his ability and mental flexibility to always ask the right questions to improve and validate my work. Next, I would like to thank my external examiner Prof. DaSilva, who was very quick in reviewing my thesis, provided a lot of useful comments, and even escaped the Queen in Dublin to come to Berlin for my defense on his birthday. Thanks Luiz!!!

Next on the list are my long-term colleagues and friends Zwilling Mati (a.k.a. Dr. Bohge Bohge) and Sven (a.k.a. Swiety) for spending all this time with me at TKN and the TU Berlin. Without you I probably would never have started my Ph.D. in the first place. To complete the group there is also Marc. Thanks guys for the fun study times (especially the learning sessions) and — not to forget — also party times. I would also like to thank Prof. Armin Zimmermann and all the Kollegiaten from the DFG Graduiertenkolleg MAGSI, who always provided very valuable feedback on my work. Thanks to Jean Bollot and Sridhar Machiraju from Sprint Labs in California especially for the DySPAN best paper. I had a really fun time with you and learned a lot. Thanks also to Ed for the support with all the administrative stuff and for the great Burrito place in San Bruno and to Jamie for the great cold-water diving in Monterey. I have to thank all the TKN folks, especially our administrative assistants Petra, Heike, and Sonja, who were always great help with everything that needed help and to our technical staff Sven, Mali, George, and Peter. Furthermore, I’d like to thank the “old” TKN folks, especially Prof. James Groß (a.k.a. Dr. Schnappi) for all the support in my early Ph.D. years, Andy (especially for the green arrow bike trail), Mama Emma and Ana (for bringing the “female touch” to the group). It was great to be able to begin my Ph.D. in such a nice atmosphere where not always only working was important. I also have to thank the “new” TKN folks for showing me that TKN can also be fun without those “old” guys. I want to especially thank Lukasz, Micha, Filip, Thomas, Niels, Danil, and Jan-Hinrich for reviewing my thesis on very short notice. Furthermore, thanks to my students Christian, Mikolaj, and Michael for lots of fun doing research and to Karl and Michael for the great work on the Mobility Framework and MiXiM.

Acknowledgments

I also want to thank all my friends who supported me, especially Bijan (my oldest friend who always has been there for me), Hannes (for many good WG parties), my chica Kirsi and Nickoh (for Sommerrollen and Wizard), Mati (again) and Frauke, and all the others I cannot name in person. Thanks also to all the guys making our San Francisco time so much fun, especially Kevin, Thomas, Marije, Afshin, Martin, Kal, Fidelma, and of course Mitch for the best SF apartment ever. Last but not least, a very special thank you to all my family, especially my parents Elisabeth and Horst and my brothers and sisters Jonas, Judith, Simon, and Rahel. Without your love and support nothing of this would have been possible. Finally, the biggest thanks has to go to my wife Sandra and my little daughter-to-come Lina. Thanks for your love, support, and understanding. And thanks to Lina for being patient enough and letting me finish my thesis before seeing the light of day. Now I can hardly await you coming.

Zusammenfassung

In den letzten Jahren wurde eine Diskrepanz zwischen den vergebenen Lizenzen für Radiospektrum und der tatsächlichen Nutzung des Spektrums beobachtet. Obwohl die überwiegende Anzahl von Frequenzbändern im für drahtlose Kommunikation attraktiven Spektrum lizenziert sind, wurde in vielen Messkampagnen gezeigt, dass große Teile des Spektrums an vielen Orten temporär ungenutzt sind. Sensing basierte opportunistische Wiederverwendung von Spektrum ist ein interessanter Ansatz um diese Diskrepanz zu beseitigen. In einem solchen Ansatz sensen sogenannte Sekundärnutzer basierend auf Cognitive Radio (CR) Technologien die Frequenzbänder von lizenzierten Primärnutzern und benutzen das entdeckte, temporär ungenutzte Spektrum für die eigene Kommunikation. Die Auflage für eine solche Nutzung ist, dass das Spektrum augenblicklich geräumt werden muss sobald ein Primärnutzer zurückkehrt. Eine offensichtliche Herausforderung für eine solche Art der Nutzung ist die zuverlässige Erkennung von Primärnutzern. Die Sekundärnutzer müssen sicherstellen, dass das lizenzierte Spektrum rechtzeitig geräumt wird wenn nötig und dass keine für den Primärnutzer schädliche Interferenz entsteht. Für solche CR Netzwerke ist allerdings nicht nur der Schutz der Primärkommunikation eine Herausforderung, sondern auch die Zusicherung eines gewissen Quality of Service (QoS) für die Sekundärkommunikation. Durch die strikte Priorität der Primärnutzer muss die Sekundärkommunikation potentiell häufig auf neue, temporär ungenutzte Frequenzen ausweichen. In dieser Arbeit präsentieren und evaluieren wir ein CR Systemdesign welches beide Herausforderungen adressiert. Wir zeigen, dass das vorgeschlagene Design sowohl einen zuverlässigen Schutz für den Primärnutzer als auch eine gewisse QoS Unterstützung für die Sekundärkommunikation erreichen kann. Das Design kann dies sogar für kleine Netzwerke mit wenigen, einfachen Sekundärnutzern, welche einen auf Energiedetektion basierenden Sensing-Prozess für die Primärnutzererkennung verwenden. Benutzt das Sekundärsystem eine gewisse Menge an spektralem Overhead für den Sensing-Prozess und die sekundäre QoS Unterstützung, kann sogar in Szenarios mit hoher Variabilität des temporär verfügbaren Spektrums eine zuverlässige Primärnutzererkennung und eine gewisse QoS Unterstützung für die sekundäre Kommunikation ermöglicht werden. Wir evaluieren den Tradeoff zwischen beiden Overheads und zeigen, dass es einen optimalen spektralen Overhead gibt, für den die spektrale Effizienz maximiert wird. Wir zeigen weiterhin, dass, obwohl die spektrale Effizienz von anfänglich kleinen Netzwerken gering ist, sie trotzdem signifikant größer ist

als Null. Des weiteren wird die spektrale Effizienz mit wachsendem Sekundärnetzwerk verbessert. Das macht unserer vorgeschlagenes CR Design zu einem idealen, skalierbaren Ansatz für initiale Netzwerkdeployments von kleinen und billigen CR Netzwerken: mit steigender Netzwerkgröße (welche normalerweise Hand in Hand geht mit einer steigenden Kapazitätsnachfrage) steigt auch die spektrale Effizienz.

Abstract

Within the last years a discrepancy between the spectrum licenses and the actual usage of spectrum has been observed. While the vast majority of frequency bands attractive for wireless communication are licensed, measurement campaigns have shown that large portions of the spectrum are temporarily unused in many locations. Sensing-based opportunistic spectrum re-usage has been identified as an attractive approach to overcome this discrepancy. In this approach Cognitive Radio (CR) based Secondary Users (SUs) sense the licensed frequency bands owned by Primary Users (PUs) for available spectrum and use the temporarily available spectrum on an opportunistic basis with the constraint to vacate the spectrum as soon as the license holder returns. An apparent challenge for such secondary spectrum usage is the reliable detection of the PU communication. The SUs have to ensure that the licensed spectrum is always vacated in a timely manner and that no harmful interference is created to the PUs. However, not only the protection of the PU communication is a challenging task for such CR networks but also the maintenance of a proper Quality of Service (QoS) for the secondary communication. Due to the strict access priority of the PUs, the secondary communication potentially has to be often relocated to new, temporarily available frequency bands. In this thesis we present and evaluate a CR system design, which is able to cope with these two challenges. We show that the proposed system design can achieve both reliable PU protection and secondary QoS support even for small secondary networks consisting of simple, low complexity CRs using energy detection-based spectrum sensing for the PU protection. Using a proper amount of spectral overhead for spectrum sensing and for secondary QoS support, reliable PU protection and secondary QoS support can also be maintained in environments with very high variability of temporarily available spectrum. We evaluate the tradeoff between both overheads and show that there exists an optimal joint spectral overhead which maximizes the spectral efficiency. We further show that, while the spectral efficiency of initially small network deployments based on the proposed system design might be low, it is still significantly greater than zero. Furthermore, the spectral efficiency is improved, as the network size increases. This makes the proposed CR design an ideal approach for initial deployments of very small and cheap CR networks, which are perfectly scalable: with an increase of network size (which usually goes hand in hand with an increase in spectrum demand), also the spectral efficiency is increased.

Contents

Acknowledgments	iii
Zusammenfassung	v
Abstract	vii
Table of Contents	ix
List of Figures	xi
List of Tables	xv
1 Introduction	1
1.1 Thesis Motivation and Goals	2
1.2 Thesis Contributions	5
1.3 Thesis Structure	5
2 Basic Concepts	7
2.1 What is a Cognitive Radio?	8
2.2 Dynamic Spectrum Access	10
2.3 Cognitive Networking	12
2.4 Regulation, Standardization, and System Designs	19
2.5 Definitions	27
3 Thesis Motivation and Approach	29
3.1 Thesis Motivation	30
3.2 Thesis Approach	31
4 Identification and Characterization of Available Spectrum	37
4.1 Spectrum Measurement Campaigns	39
4.2 Limitations of the Presented Spectrum Measurement Campaigns	45
4.3 Ground Truth of Spectrum Availability	50
4.4 Estimating the Variability of Available Spectrum	57

5	Reliable Detection of Primary Users	59
5.1	Performance Metrics	60
5.2	Is Sensing-based Detection of Primary Users Possible?	61
5.3	Achievable Diversity in Spectrum Sensing	70
5.4	Related Work	77
5.5	Conclusions	78
6	Cognitive Radio System Design	81
6.1	Secondary User Links	82
6.2	Periodic Spectrum Sensing	83
6.3	Link Reconfiguration	86
6.4	Secondary QoS Maintenance	88
6.5	Feasibility Assessment	91
7	Secondary Quality of Service Support	93
7.1	Performance Metrics	94
7.2	Impact of Primary User Agility and False Positives on Secondary QoS	95
7.3	QoS Support in Dynamic Frequency Hopping Systems	104
7.4	Related Work	106
7.5	Conclusions	107
8	Spectral Efficiency in Cognitive Radio Networks	109
8.1	Performance Metrics	110
8.2	Spectrum Sensing and Secondary QoS Support Efficiency Tradeoff	111
8.3	Spectral Efficiency in Dynamic Frequency Hopping Systems	120
8.4	Related Work	122
8.5	Conclusions	123
9	Conclusions and Outlook	125
9.1	Conclusions	126
9.2	Future Work	127
A	Modeling Cellular Primary Usage	129
B	Reliable Detection of Cellular Primary Users	137
C	Acronyms	159
D	Variables and Symbols	163
E	Publications	167
	References	171

List of Figures

2.1	Examples for contiguous and non-contiguous Secondary User Links	18
4.1	Electromagnetic spectrum	38
4.2	Normalized load of three different cells over 3 weeks.	52
4.3	Distribution of system-wide average call arrival rates.	53
4.4	Distribution of average call duration over 5-minute periods.	54
4.5	Average per-cell variation of load on an hour-of-day basis.	55
4.6	Maximum change in load plotted on an hourly basis	55
4.7	Maximum change in load plotted as a function of the maximum interference time	56
5.1	Single sensor Receiver Operating Characteristic	67
5.2	Multi sensor Receiver Operating Characteristic	68
5.3	Frequency diversity in measurement 1	72
5.4	Comparison of the frequency diversity in the different measurements . . .	72
5.5	Time diversity in measurement 1	74
5.6	Comparison of the time diversity in the different measurements	74
5.7	Diversity tradeoff for TV bands	76
5.8	Diversity tradeoff for cellular bands	77
6.1	Cognitive Radio system bandwidths	82
6.2	Example distributions of sub-channels for a SUL with bandwidth $b_{sul} = 4$ sub-carriers.	83
6.3	Approaches for periodic sensing organization	84
6.4	Time domain sensing model	85
6.5	Redundancy example	89
6.6	Resource reservation example	90
7.1	Secondary QoS support system model	95
7.2	Impact of the PU system load and the PU activity pattern on the average link capacity	99
7.3	Impact of the number of sub-channels of a SUL on the average link capacity	100

7.4	Impact of the slot length and link reconfiguration time on the average link capacity	101
7.5	The α quantile of the link capacity for $\alpha = 0.9999$ (left column) and $\alpha = 0.95$ (right column)	102
7.6	Impact of the probability of false positives on the average link capacity	103
7.7	Impact of the actual PU system load and the link reconfiguration time on the average link capacity	104
7.8	Influence of the probability of false positives on the α quantile of the link capacity	105
7.9	Secondary QoS support system model for Dynamic Frequency Hopping	106
8.1	Comparison of the spectrum sensing and the QoS maintenance spectral overhead	117
8.2	Investigation of the optimal sensing bandwidth	118
8.3	Influence of operation environment and QoS requirement on the effectively available spectrum	119
8.4	The effectively available spectrum for the DFH approach.	122
A.1	Histogram of call durations	130
A.2	Duration distributions and lognormal fits.	131
A.3	Illustration of anomalous distributions during 2 hours.	132
A.4	The percentage of successful fits (across all cells)	133
A.5	The per-hour auto-correlation of the step sizes (Φ)	134
B.1	Measurement locations (in red) and nearest cell tower (in blue).	139
B.2	Dataset <i>D1</i> : Sensed power over time for carriers <i>C1</i> and <i>C2</i>	141
B.3	Dataset <i>D1</i> : Normalized load for strongest cell sectors.	141
B.4	Cross-correlation and time synchronization of sensed power and ground truth load	142
B.5	Dataset <i>D1</i> : Distribution of power levels when the load varies.	143
B.6	An illustration of how power changes when various call events happen	144
B.7	Dataset <i>D1</i> : 10-second <i>t-average</i> plots for (downlink) power during call initiation and termination events.	144
B.8	Dataset <i>D2</i> : 10-second <i>t-average</i> plots for (downlink) power during call initiation termination events.	145
B.9	Dataset <i>D1</i> : 10-second <i>t-average</i> plots for (uplink) call initiation and termination events	146
B.10	Dataset <i>D1</i> : cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 1	148
B.11	Dataset <i>D1</i> : cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 2	149

B.12 Dataset *D1*: cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 3 150

B.13 Dataset *D1*: Fraction of identified initiation events and false positives by the detectors based on the 3 criteria with varying threshold values. 151

B.14 Dataset *D1*: cCDF plot for call termination. 152

B.15 Dataset *D1*: Call arrival and load overestimation 152

B.16 Dataset *D1*: (Over)estimation of unused capacity. 153

B.17 Dataset *D1*: Probability of estimating at least X % of unused capacity. . . 154

List of Tables

2.1	Typical standard deviation of the delay spread (σ_{ds}) for different propagation environments	15
2.2	(Over)DRiVE	23
2.3	DARPA XG	24
2.4	Spectrum Pooling	25
2.5	CORVUS	25
2.6	Spectrum Agile Radio	26
2.7	DIMSUMnet	26
2.8	IEEE 802.22	27
4.1	Spectrum Occupancy in New Zealand	41
4.2	Spectrum Occupancy Table 30 – 960 MHz	42
4.3	Spectrum Occupancy Table 960 – 3000 MHz	43
7.1	Important parameters and values used for the secondary QoS evaluation	98
8.1	Important parameters for the calculation of the spectrum sensing spectral overhead.	112
8.2	Important parameters for the calculation of the QoS maintenance spectral overhead.	114
8.3	Important parameters and values for the efficiency tradeoff evaluation	116

CHAPTER 1

Introduction

1.1 Thesis Motivation and Goals

Wireless spectrum is a scarce resource. Especially the lower frequency ranges (up to 6 GHz), which are very attractive for wireless communication due to their good propagation characteristics, are completely allocated by the different national regulation authorities. The UMTS license auctions in Europe have shown how much money telecommunication providers are willing to pay for licenses in these attractive frequency bands. The actual *usage* of the wireless spectrum, however, demonstrates the waste of spectrum due to the fixed allocation. Many measurement campaigns have shown that large portions of the spectrum are *temporarily* unused in *certain* locations.

An interesting way to overcome this discrepancy between spectrum allocation and spectrum usage is to allow so called *secondary usage of the spectrum* applying Dynamic Spectrum Access (DSA) technologies. Within this concept the license holders of the spectrum have strict priority in access to the spectrum. Within this thesis we refer to such license holders as Primary Users (PUs) and to the spectrum band owned by a Primary User as PU band. Secondary Users (SUs) can use spectrum of a PU band only if the respective PU is not present, with the constraint that they have to vacate the spectrum again as soon as the PU returns. In such a case, SUs will usually look for free spectrum in other PU bands to continue their communication. A Secondary User Group (SUG) is defined as a group of SUs (terminals or a base station with associated terminals) sharing a communication channel in order to support mutual exchange of information. Several SUGs can share the same spectrum (similar to today's Wireless Local Area Networks (WLANs)). In general, each SUG individually decides if the PU re-appeared and similarly can make its own decision regarding which spectrum band to switch to.

Many approaches are being discussed in the research community about how to realize such secondary spectrum usage. The PU could simply announce unused spectrum, which then could be used by the SUGs, or auctions could be organized to make temporarily unused spectrum available for SUGs. The advantages of such approaches are that non-interference to PUs can be guaranteed, and that the regulation of the secondary spectrum access is easy and has low complexity. On the other hand, these approaches have the drawback that the PU systems have to be enhanced (e.g., for announcing / auctioning free spectrum) or even modified (e.g., once spectrum has been auctioned for a certain time span, it cannot be used by the PU during that time).

An alternative, very attractive approach for the realization of secondary spectrum usage is opportunistic spectrum sharing based on spectrum sensing. In this approach, the identification of idle spectral resources as well as the detection of PUs appearing on spectrum currently used for secondary communication is based on a spectrum sensing process performed by the SUGs (or a dedicated spectrum sensing infrastructure). The big advantage of this approach is that the secondary usage is (ideally) unrecognizable to the PUs and no changes need to be made to the licensed systems. The PU systems can operate as if there would be no secondary usage of their PU bands. The drawback is a very high complexity of the secondary system. PUs have to be detected autonomously,

and regulatory rules are required to specify under which circumstances a PU band can be assumed to be unused. Furthermore, sensing-based detection unavoidably results in (very short) interference to the PU systems. It is again the task of regulation to specify the limits for which such interference is considered to be harmful to the PUs.

Crucial for such an approach is to develop reconfigurable and “intelligent” radios which are able to adapt to their environment — a concept pioneered by Mitola III [115] and referred to as Cognitive Radio (CR). The idea of CR-based secondary spectrum usage has been very intensively discussed worldwide. It got an additional boost due to the fact that the U.S. regulator FCC has defined rules for such spectrum usage: the utilization of unused channels (so called “white spaces”) in the TV bands [47] with the main goal to provide wireless broadband access in rural areas.

Cognitive Radio is a rather young research area and there are still a lot of challenges and unsolved issues. One of the main challenges is the reliable detection of PUs, which is an enabling functionality for sensing-based secondary spectrum usage. To facilitate the detection process in the case of the white space usage in the TV bands, the FCC requires a geo-location and database-assisted PU detection process. This means that a SUG has to query a database maintained by the PUs and regulation authorities, indicating the PU activity. Originally, spectrum sensing was additionally required for devices to be operated in the TV bands. However, in a very recent ruling from September 2010, the FCC eliminated the need of spectrum sensing in addition to the geo-location database-assisted detection [49], making secondary TV white space usage into a purely announced sharing approach.

One aspect enabling such an approach in the TV bands is the rather static behavior of the PUs. If a TV station occupies its spectrum, it usually does so continuously for weeks or months, which also implies that a SUG, once it finds spectrum temporarily not used by a PU, will likely be able to use it for a long time without the PU returning. Another aspect contributing to the static operation environment is the stationarity of the envisioned CR system consisting of a stationary base station and so called Customer Premise Equipments (CPEs), which are installed in people’s homes and, thus, also stationary.

The FCC rulings for secondary usage of white space in the TV bands were the driving factor for the creation of the IEEE 802.22 working group [73] which develops standards for the operation of Wireless Regional Area Networks (WRANs) in the TV bands on a license-exempt, non-interfering basis. These efforts will soon result in the first deployments of CR-based systems for white space usage in the TV bands, which is an important step to show the realizability of such approaches and will provide lots of insights for future research. More advanced systems for secondary usage, however, also should be able to cope with challenging scenarios such as more agile PUs or mobility of the PU and / or CR network. Note that, to a SUG, mobility results in a *perceived* increase of PU agility. For such scenarios database assisted PU detection will hardly be feasible due to the very dynamic spectrum availability, which can change rapidly and often. The corresponding challenging research question to be addressed in this thesis is: **“How can sensing-based secondary spectrum usage be realized in scenarios**

where **SUGs have to deal with high dynamics of spectrum availability** (due to agile PUs and / or due to mobility of the PUs / SUGs)?”

Another aspect to consider is ease (and costs) of deploying CR-based systems. The need to deploy and maintain databases or to deploy a sensing infrastructure to help in the detection of licensed users may be a show-stopper for CR systems to be rolled out. Furthermore, the requirement of high precision, complex CR front-ends for spectrum sensing and DSA can significantly increase the time-to-market of initial CR deployments. Innovative secondary usage approaches should take these aspects into consideration. These requirements can be formulated as a second challenging research question to be addressed in this thesis: “**How to enable secondary spectrum usage with simple, low cost devices without blocking the possibility to gradually enhance the system and its performance?**”

Summarizing, the focus of this thesis is on sensing-based opportunistic spectrum sharing. The main question to be answered is: “**How can sensing-based secondary spectrum usage be realized in environments with very dynamic spectrum availability with the constraint of using simple, low cost CRs, which can be easily deployed?**”

To assess the above research question and evaluate the performance of candidate systems for such a challenging secondary usage, a triangle of three fundamental performance metrics is formulated.

The most important enabling functionality for secondary spectrum usage is the **reliable protection of the PU communication**. It has to be ensured that SUGs only utilize PU bands for their communication needs which are currently not used by any PU. Furthermore, the quick and reliable vacation of PU bands reclaimed by a PU has to be ensured.

Recognizing the requirement of reliable PU protection, DSA and secondary usage of spectrum can only become a service reality if at least some degree of **Quality of Service (QoS) support for the secondary communication** is provided, which is the second performance metric. Due to the requirement to always avoid PUs, secondary communication potentially often has to be relocated to different, currently available spectrum. Thus, the fundamental secondary QoS metric is to which extent service interruptions can be avoided and / or compensated for.

The third performance metric to be considered is the **spectral efficiency of the secondary usage**. The driving force for DSA and secondary usage of spectrum is the inefficient spectrum utilization of legacy systems. Whereas it could be argued that any additional usage improves the spectral utilization, innovative DSA approaches should be designed with the goal of efficient spectrum usage in mind. In particular, spectral efficiency becomes critical as soon as the demand for spectrum is bigger than its availability.

Having defined the metrics above, it is important to understand that reliable protection of the PU communication is a strict requirement which has to be fulfilled. The other two metrics, in contrast, are flexible, allowing for different system design approaches, and can be used to evaluate the performance of a secondary system.

1.2 Thesis Contributions

The first main contribution of this thesis is an investigation of the variability of the spectrum available for secondary communication. It is shown that not only the agility of primary usage impacts the variability of available spectrum, but also mobility of the primary and secondary systems and the quality (false positives) of the sensing process. To quantify the agility of primary usage, a detailed study of the spectrum occupancy of cellular PUs (voice calls) is presented. The study is based on measurements inside the network of a big U.S. network provider, and results in a model for the spectrum usage of cellular PUs.

The second main contribution is an evaluation of the spectrum sensing based PU detection process. The requirements for the sensing process being able to reliably detect PUs in very dynamic spectrum environments are determined and evaluated. Furthermore, the influence of the different diversity approaches for spectrum sensing (and combinations thereof) on the PU detection process is investigated. The results show that reliable PU protection is possible also for very agile PU usage characteristics given a certain amount of diversity and spectral overhead.

Based on the result that reliable PU detection is possible, the third main contribution is the proposal of a CR system design, which is able to provide a certain secondary QoS while fulfilling the PU protection requirement. The proposed design is based on innovative approaches that reduce the impact of potentially frequent reconfigurations of the secondary communication due to spectrum sensing and the variability of available spectrum. To further improve the secondary QoS support, additional mechanisms to compensate for the service interruptions of the secondary communication due to reconfigurations are proposed.

For the proposed system design, a simulation-based evaluation of the impact of selected system parameters on the QoS of the secondary communication is presented, which is the fourth main contribution. It is shown that the proposed system design can provide a reasonable secondary QoS for very diverse dynamics of available spectrum including scenarios with a very high variability of available spectrum.

Additionally, as the fifth main contribution, a simulation-based investigation of the spectral efficiency of the above presented CR system design is presented. The spectral efficiency of the sensing process as well as the secondary QoS support mechanisms are analyzed and the impact of, and the tradeoff between, both are quantified. It is shown that for the proposed system design the spectral efficiency improves as the number of CRs in the system increases.

1.3 Thesis Structure

In the following the general structure of this work is outlined to provide an overview of the contents of the presented thesis.

In Chapter 2 the basic concepts of Cognitive Radio (CR) and Dynamic Spectrum Access (DSA) are introduced. While some introduction is given to the general concepts,

the focus is put on cognitive networking aspects important for the remainder of the thesis. This chapter is mainly intended for the reader not familiar with the presented topics; however, it also introduces some definitions used throughout the thesis.

The detailed motivation and approach taken in this thesis are presented in Chapter 3. The chapter gives an outline of the structure and approach for this thesis.

Chapter 4 summarizes and evaluates spectrum measurement campaigns of the last years analyzing the spectrum occupancy of licensed systems. A special focus is put on the variability of spectrum occupancy. Based on a big measurement campaign investigating the real spectrum occupancy as recorded inside a primary network, a PU occupancy model for the cellular bands is presented. Furthermore, it is shown that the agility of spectrum usage by PUs is not the only influencing factor for the variability of available spectrum.

Reliable detection of PUs is especially challenging if operating in environments with very dynamic spectrum availability. In Chapter 5 a feasibility study is presented of whether and under which circumstances reliable detection in such challenging environments is possible. Furthermore the achievable diversity is investigated for selected scenarios.

In Chapter 6 a CR system design is presented, which is able to reduce the influence of variability in available spectrum on the secondary QoS while at the same time meeting the PU protection requirement. Furthermore, QoS compensation mechanisms are presented for the proposed system design.

Secondary QoS support is a very important metric for cognitive networking. The requirement to quickly vacate spectrum where a PU appeared makes secondary QoS support especially challenging in case of very dynamic spectrum availability. In Chapter 7 the impact of PU agility and sensing accuracy on the secondary QoS for the presented system design is investigated.

Since secondary spectrum usage is a means to overcome the underutilization of the spectrum, CR systems should be designed with the goal of spectral efficiency in mind. Both spectrum sensing and secondary QoS support require spectral overhead and the tradeoff between both overheads is investigated in Chapter 8. Furthermore, the previously presented CR system design is investigated with respect to its spectral efficiency.

In Chapter 9 the major contributions of the work presented in this thesis are summarized and general conclusions are drawn. Furthermore, pointers towards potential future work are given.

CHAPTER 2

Basic Concepts

In this chapter the basic concepts of Cognitive Radio (CR) and Dynamic Spectrum Access (DSA) are presented. Furthermore, the basic building blocks for cognitive networking are introduced and an overview of proposed system designs and relevant related work is given. In the last section of this chapter the main definitions used within the thesis are summarized.

2.1 What is a Cognitive Radio?

There are many different terms and definitions which are used within the research community in the context of cognitive radio. The most commonly used terms are “Cognitive Radio (CR)”, “frequency agile radio”, “software radio”, and “Software Defined Radio (SDR)”. These terms are often used as synonyms; however, they also might be used to emphasize specific features.

This section summarizes some commonly used definitions and introduces the terminology and definitions used in this thesis. Furthermore, the most important concepts of a CR for this thesis are introduced.

2.1.1 Cognitive Radio Definition

The term Cognitive Radio (CR) was originally introduced by Mitola III who suggests the following definition [112]:

“The term cognitive radio identifies the point in which wireless personal digital assistants (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer-to-computer communications to:

- (a) detect user communications needs as a function of use context, and
- (b) to provide radio resources and wireless services most appropriate to those needs.”

Mitola III, thus, defined Cognitive Radio (CR) in a very broad sense, not only including radio spectrum and wireless communication aspects but additionally including the ability to assess the needs and requirements of a user. The basic idea of Mitola III is to have computer devices which can automatically assess and predict the communication needs of a user and always provide the best (and / or cheapest) required connectivity. Within the wireless communication community the Mitola III definition is usually reduced to the radio communication part, i.e. to part (b) of the definition.

Mitola III introduced CR as an extension to software radios [116], originally introduced in [114]. Software radios, often also referred to as Software Defined Radios (SDRs), denote radios where functionality which was usually realized in hardware is implemented in software. The advantage of this is the possibility to easily change radio parameters on the fly, making the radio extremely flexible. Nowadays, the term Cognitive Radio is mostly used in a slightly broader sense, not limiting CRs to software radios. The FCC [46] suggests that any radio having adaptive spectrum awareness should be referred to as a CR. More precisely:

“A cognitive radio (CR) is a radio that can change its transmitter parameters based on interaction with the environment in which it operates. The majority of cognitive radios will probably be SDRs (Software Defined Radios), but neither having software nor being field programmable are requirements of a cognitive radio.”

Note that, according to the above definition, even IEEE 802.11 access points implementing proprietary automatic channel selection mechanisms can be considered as very basic CRs. The FCC definition also supports the differentiation between CR and “frequency agile radio” on the one hand, and “software radio” or SDR on the other hand. Whereas the first two terms describe the ability of a radio, the focus of the second two terms is on how certain functionalities are realized. Throughout this thesis, the FCC definition of CR is followed.

2.1.2 Main Cognitive Radio Functionalities

Based on the FCC [46] definition, the focus within this thesis is on the communication-specific functionalities of a CR, i.e., the ability to provide means of communication based on the user’s needs. The prediction of the user’s communication needs and the learning based on the user behavior, as specified in part (a) of the definition by Mitola III [112] are not considered.

Based on this, there are two main functionalities a CR needs to provide: detecting available spectrum and adapting its transmission to the available spectrum. In opportunistic spectrum sharing, detection of available spectrum is done *autonomously* by a CR (or a group of CRs) through a sensing process. Having identified the currently available spectrum, the CRs *autonomously* decide which parts of the available spectrum to use and how to use it. If spectrum used by the CRs becomes occupied by a Primary User (PU), this has to be detected by the sensing process and the transmission parameters have to be adapted accordingly.

The technical details of how to design radios to be able to perform the spectrum sensing process and to support the required flexibility to adapt the secondary transmission parameters (such as center frequency, transmission bandwidth, transmission technology, etc.) based on the currently available spectrum are beyond the scope of the presented work. However, there are some general radio design concepts which are important to understand and, thus, summarized here. The most basic CR consists of one radio front-end which is used for both spectrum sensing and as a transceiver, but can only perform one of these tasks at a time. During the spectrum sensing process, no data transmission can take place and vice versa. More advanced CRs can perform both tasks in parallel: while performing sensing in one part of the spectrum, data transmission can be performed in another part of the spectrum. Note that whether this is achieved by multiple radio front-ends or by one, more complex, radio front-end is conceptually irrelevant. Another distinguishing feature for CRs is whether the transceiver can operate in non-contiguous spectrum. Today, almost all commercial wireless radio technologies are only able to transmit using one single channel (a continuous chunk of spectrum) at a time. Whereas this channel can often be shifted within the operation range of the radio, it is not possible to communicate using multiple non-contiguous channels at the same time. State of the art wireless research prototypes based on Software Defined Radio (SDR) concepts, in contrast, often have the ability to operate using non-contiguous channels. For CRs concurrently operating in the PU bands of multiple PUs, this is an essential feature.

2.1.3 Policy Driven Operation

Another important concept of the used definition of CR is policy driven operation. As CRs work autonomously, policies are an important concept to regulate them. The idea of a policy driven operation is that a CR can dynamically load and apply different policies that regulate its behavior depending, e.g. on the current location.

Policies will usually be defined by regulation or by the PUs itself. They define the rules for secondary usage of spectrum. Before starting to operate in some licensed PU band, a CR needs to acquire the respective policies for that PU band. Note that policies might differ depending on location (e.g., for different countries) and that policies also might expire (e.g., due to changes in regulation). A CR, thus, has to ensure to have an up-to-date set of policies for the region it is operating in.

A very important aspect is that CRs are able to execute the policies. Thus policies have to be defined in a machine understandable language. The DARPA neXt Generation (XG) program: [34] specifies the use of a policy language [36] for the XG radio system [108]. In [13] details of the proposed policy language are explained. The concept of a policy language for CR is further discussed in [113, 117].

As mentioned above, CRs have to somehow get the policies and updates of the policies. Since policies define the behavior of CRs, authentication and integrity of policies is very important. Some of these aspects are also discussed in the XG context. A good overview on policy languages and management of policies can be found in [125, 126].

2.2 Dynamic Spectrum Access

The idea of Dynamic Spectrum Access (DSA) is to continuously adapt the radio's own spectrum usage to the current spectrum occupation of other users or technologies. The motivation for DSA is to overcome two major problems of the static spectrum access of most of today's wireless technologies: spectrum underutilization in the licensed spectrum bands and interference in the unlicensed spectrum bands (and among different secondary networks also in the licensed spectrum bands).

Within the unlicensed "public commons" (Unlicensed National Information Infrastructure (UNII) and Industrial Scientific and Medical (ISM) bands) the spectrum is shared between many users and, in general, many different technologies. Bluetooth, WLAN (IEEE 802.11), sensor networks (IEEE 802.15.4), cordless phones, and wireless garage openers are only some technologies operating in the very popular 2.4 GHz ISM band. The predominantly static spectrum access of these technologies creates more and more interference as the number of users and technologies increases, resulting in collisions and severe service degradation of the individual technologies. By applying DSA, interference can be reduced and the available spectrum can be used more efficiently. In fact, due to the rapidly increasing spectrum usage in the 2.4 GHz ISM band, many modern WLAN access points already support a (proprietary) dynamic channel selection feature, automatically selecting the WLAN channel with the least interference. This is one of the examples of CR functionality without the usage of Software Defined Radio (SDR)

technology.

In the context of licensed spectrum, DSA can help to overcome the underutilization of the spectrum by the PUs. CRs can use DSA techniques to fill the gaps created by the static spectrum access of the PUs and utilize the temporarily unused spectrum. The focus of this thesis is on DSA as a means for sharing spectrum between PUs and CRs. There are three different approaches to such spectrum sharing: opportunistic, negotiated, and announced spectrum sharing. In opportunistic spectrum sharing CRs *coexist* with PUs, whereas in negotiated and announced spectrum sharing CRs *cooperate* with PUs [122].

2.2.1 Announced Spectrum Sharing

The idea of announced spectrum sharing is that the PU (or a broker working on its behalf) announces both the availability of free spectrum and the time period for which the availability will hold. This announcement is understood as a clear release of the spectrum by the PUs for the announced time period. A CR has to listen for announcements and possibly has to compete with other CRs for the idle spectrum resources announced by the PU.

If the announcements are used properly, no interference is caused to the PU. On the other hand, announced sharing requires the cooperation of PUs, which have to declare the spectrum release. Note that no QoS guarantees for the secondary communication can be given, since the idle spectrum is accessed in a non-coordinated, competitive manner by the CRs. The emerging IEEE 802.22 standard [82] follows the announced spectrum sharing approach. CRs have to consult a database specifying the activity on the different TV channels. The additional requirement of spectrum sensing has been eliminated by the most recent FCC ruling [49].

2.2.2 Negotiated Spectrum Sharing

In negotiated spectrum sharing a given CR can only use the spectrum with explicit permission (e.g. from the license holder or a spectrum broker). A CR, thus, has to explicitly request usage of some spectrum for a certain time period. This can obviously be based on analysis of announcements as discussed in the previous case. The permission to use the spectrum is given for some limited time.

In negotiated spectrum sharing the secondary system does not create any interference to the legacy system. Additionally, even QoS guarantees for the secondary communication can be temporarily given, as a CR can, e.g., negotiate to use a certain spectrum range exclusively for a certain time. Alternatively, some CRs might get authorization to use a chunk of spectrum jointly according to a specific sharing mechanism. In addition to the announcements of available spectrum, the usage of spectrum is managed directly by the PUs or by a third party (e.g., a broker) on behalf of the PUs. An example of negotiated spectrum sharing is DIMSUMNet [18, 19], where a spectrum broker leases spectrum to the CRs. Possible options are also spectrum auctions [84, 143].

2.2.3 Opportunistic Spectrum Sharing

In opportunistic spectrum sharing, spectrum access is based on identifying the presence of PUs by a sensing process. The rules and conditions for this sensing process as well as the usage of the spectrum are defined by (regulatory) policies. The sensing is mostly done by the CR itself; however, sensing can also be done (or complemented) by external sensing devices specifically deployed by third parties (e.g. regulators or the PU itself).

An important condition for opportunistic spectrum sharing is the requirement that a CR must not prevent the PU from accessing its PU band. For PUs using a Carrier Sense Multiple Access (CSMA)-based access scheme, this implies, e.g., that CRs always have to operate below the sensitivity threshold of the PU. More generally, this means that, within a PU band currently used for their secondary communication, CRs must be able to detect a PU which (re)starts usage of its spectrum. This detection is usually done by the sensing process and is quantified by the probability of false negatives (P_{fn}), i.e., the probability that a PU is not detected by the CR system within a certain time.

This time, referred to as maximum interference time (t_{max}), is defined as the maximum amount of time for which a certain PU, resuming its activity after a break in spectrum usage, can tolerate interference from secondary usage. The CR system has to ensure that the PU band is vacated within t_{max} after a PU appears in its PU band. Obviously, t_{max} puts a strict upper limit on the sensing time as well as the time a PU band can be used without being sensed. For details on sensing-based PU detection refer to Section 2.3.1 below.

This concept leads to a highly flexible and dynamic secondary usage where CRs can adapt to the local behavior of the PU. The PUs are not aware of the secondary usage of their spectrum. In fact no changes have to be made to legacy systems working as PU. Obviously, opportunistic spectrum sharing highly depends on the quality of the sensing process. Examples of system designs following the opportunistic spectrum access idea are CORVUS [17, 20] and spectrum pooling [155]. A recent prototype implementation [12] of opportunistic spectrum sharing and the testing of industrial prototypes for *sensing-based* TV band white space detection [48] shows the growing interest of industry in DSA.

2.3 Cognitive Networking

Having understood the basic concepts of a Cognitive Radio (CR) and the different approaches to secondary spectrum usage, the core functionalities necessary to realize secondary spectrum usage are explained in this section. The focus throughout this thesis is on sensing-based, opportunistic spectrum sharing, being the most flexible solution.

To utilize the temporarily unused spectrum of PUs, at least two CRs need to “network” and establish a communication link. More generally, a set of CRs (terminals or a base station with associated terminals) form a Secondary User Group (SUG) and jointly use temporarily unused spectrum for own communication purposes. Spectrum used by a SUG is referred to as the Secondary User Link (SUL) and assumed to be used in a shared modus to support communication among all the CRs involved.

The main functionalities of an individual CR identified in Section 2.1.2 build the foundation to enable cognitive networking within a SUG as explained in the following sub-sections.

2.3.1 Spectrum Sensing

Spectrum sensing is the process of deciding — based on spectrum measurements — whether or not a PU is present in its PU band. The measurement techniques used range from simple energy detection to matched filter or cyclo-stationary feature detection [148]. An extensive survey on sensing algorithms can be found in [176]. Conceptually, two different sensing processes need to be performed: *initial* sensing and *periodic* sensing.

Initial sensing has to be performed before any spectrum will be assigned for secondary usage. The initial sensing process often covers a wide spectrum range trying to determine which PU bands — possibly strongly differing in their features — are available for secondary communication. It is usually not very time critical.

Out of the spectrum detected to be available, a SUG forms a Secondary User Link (SUL), which is jointly used for communication within the group. After the setup of a SUL, *periodic* sensing has to be done regularly during the whole period of its usage. The SUG has to verify that the used resources are still available, i.e., that no PU has appeared in PU bands used by the SUL. In contrast to the initial sensing process, the periodic sensing process can be constrained to spectrum resources used by the SUL. Periodic sensing is time critical since a PU band reclaimed by a PU has to be vacated within the maximum interference time (t_{\max}), as specified in Section 2.2.3, which puts a strict upper limit on the sensing time.

A fundamental problem of spectrum sensing can be stated as follows. It is (almost) impossible to reliably detect the PU in a certain spectrum range, while at the same time performing data transmission in that range. It is, thus, crucial to ensure that the spectrum range to be sensed is not used for data transmission by the SUG at the same time. Recently, there has been work on single channel, full duplex wireless communication [28], which could potentially be used to parallelize spectrum sensing and data transmission. For the work presented in this thesis, however, we focus on the case where parallel spectrum sensing and data transmission in the same spectrum is not possible.

Performance Metrics

The fundamental performance metric of spectrum sensing is *reliability*, i.e., to ensure that interference to PUs is below a very strict threshold tolerable by the PUs. The usually applied measure for reliability is the probability of false negatives (P_{fn}), i.e., the probability of not detecting a PU in a pre-defined time period although it is present. To ensure the protection of the PU, the sensing process has to ensure to keep P_{fn} below a certain threshold (usually defined by regulation). The probability of detecting a PU although it is not present (probability of false positives (P_{fp})) is a good additional performance measure for the sensing process. We refer to the latter metric as the *quality* of the sensing process.

Meeting the requirement of reliability (false negatives) and achieving a good quality of the sensing process (false positives) comes at a cost: the *complexity* and the *overhead* of spectrum sensing. Primarily the *spectral overhead* will be considered in this thesis. Given the constraint that no data transmission can take place within the spectrum being sensed, the amount of spectrum used for sensing (defined as the product of the spectrum bands used for sensing and the time in which these bands are sensed) results in spectral overhead. In addition, the exchange of sensing information between the SUG members also requires spectral resources. Apart from the spectral overhead, spectrum sensing costs are also influenced by the sensing hardware and software complexity, the amount of energy expended for sensing, etc. For an overview of the possible overheads of spectrum sensing refer to [65].

Diversity Approaches

Diversity of received radio signals is caused by the stochastic nature of the wireless channel. On its way from the transmitter to the receiver, the signal is attenuated, reflected, diffracted, refracted, scattered, and absorbed. These alternations of the signal are usually differentiated into three effects: path-loss, shadowing, and fading. A good tutorial-style overview of these effects and of models used to describe them can be found in [3]. Fading has the biggest influence on the diversity of radio signals. Generally speaking, fading describes the frequency- and time-selective attenuation of radio signals due to multi-path propagation and reception. The main cause for diversity of received signals are the frequency offset due to the *Doppler spread* caused by movements of transmitter, receiver, or objects within the signal propagation path between transmitter and receiver, and the *delay spread* caused by the time delay with which different (reflected, scattered, etc.) copies of the same signal arrive at the receiver.

Due to this diversity of radio signals “one short sensing sample” does not provide reliable information on the presence of a PU. While the signal received by the spectrum sensor might be strong enough to be detected as a PU signal at a specific point in time, it might be too weak to be detected (due to fading) only shortly after. The same applies to the frequency and spatial domain: whereas the signal might be strong enough to be detected in some part of the PU band (at some location) it might be too weak in a different part of the PU band (at a location only centimeters away). To be able to quantify the diversity of radio signals we first introduce definitions for coherence time (t_{coh}) and coherence bandwidth (b_{coh}) as metrics for time and frequency diversity, respectively. We furthermore give an intuition on the scope of spatial diversity. In wireless communications, coherence is the inverse to diversity. Two signals are said to be coherent if they are strongly correlated, whereas two signals are diverse if their correlation is negligible. Due to historical reasons, usually the terms coherence time and coherence bandwidth are used in the time and frequency domain, respectively, whereas spatial diversity is used for the spatial domain. The incentive behind this is that in the time and frequency domain usually coherence is a desired property of radio signals to achieve a high capacity, whereas — especially in the context of Multiple Input Multiple Output (MIMO) systems — spatial diversity of transmit / receive antennas is used to

Table 2.1: Typical standard deviation of the delay spread (σ_{ds}) for different propagation environments

Environment	σ_{ds}
Urban	1 – 25 μs
Suburban	0.2 – 2 μs
Indoor	25 – 250 ns

increase the capacity.

The coherence time (t_{coh}) is mainly influenced by the Doppler spread [3]. The maximum Doppler shift (f_{d}) can be approximated using the carrier frequency (f_{c}) and the average velocity (v) of transmitter, receiver, and any objects within the path between transmitter and receiver [131] as shown in Equation (2.1).

$$f_{\text{d}} = \frac{v \cdot f_{\text{c}}}{c} \quad (2.1)$$

Using the maximum Doppler shift, one possibility to estimate the coherence time (t_{coh}), as used in [23], is shown in Equation (2.2). Other estimations can, e.g., be found in [129, 131, 141].

$$t_{\text{coh}} = \frac{1}{\sqrt{2\pi} f_{\text{d}}} \quad (2.2)$$

The delay spread is the main influencing factor for the coherence bandwidth (b_{coh}). It is mainly determined by the topology of the propagation environment. Typical values for the standard deviation of the delay spread (σ_{ds}), taken from [3], can be found in Table 2.1. The coherence bandwidth (b_{coh}) can be estimated using the standard deviation of the delay spread. As with the coherence time, there are many different proposals for estimating b_{coh} [23, 129, 131, 141]. We follow the proposal of [23], where the coherence bandwidth is estimated to

$$b_{\text{coh}} = \frac{1}{2\pi \sigma_{\text{ds}}} . \quad (2.3)$$

Spatial diversity has so far been mainly investigated in the context of MIMO systems. For those systems it has been shown that even signals received by antennas only a few wavelengths apart are largely uncorrelated. For an overview of spatial diversity in MIMO systems, refer to [39]. For spectrum sensing performed by different CRs this means that, even if there are very close together, the received signals can assumed to be uncorrelated.

Throughout this thesis we follow the frequently used block fading model [107]: signals received within the coherence bandwidth and coherence time are assumed to be coherent whereas they are assumed to be independent and identically distributed (i.i.d.) beyond.¹

¹Note that while this is a commonly made assumption, in reality multiple coherence times (or coherence bandwidths) might be required to get i.i.d. signals. Coherence means strongly correlated, so signals being one coherence time apart are likely still at least correlated to some degree. However, the results presented in this thesis would only change quantitatively if using a more strict assumption regarding the independence of signals that occur beyond the coherence time / coherence bandwidth.

Signals received by two different sensors in different locations (spatial diversity) are always assumed to be i.i.d. A single *sensing sample* is obtained by sampling the received radio signal within a frequency span of b_{coh} and for a time of t_{coh} . Note that the sampling frequency (f_s) used to sample the radio signal is usually different from the frequency with which sensing samples are recorded. The received signal is usually sampled at Nyquist rate and many of these micro-samples are used to obtain one *sensing sample*.

To combat the effects of fading and to increase the accuracy of the sensing process, the diversity of i.i.d. sensing samples can be exploited. To do so, the sensing samples are combined to achieve a decision based on all samples. There are two different concepts for the merging of sensing samples: merging of raw measurement data (soft decision combining) or merging of 1-bit decisions (hard decision combining) [152]. For hard decision combining, each sensing sample is used to compute a one bit decision (PU present or not) and these one bit decisions are merged to a joint result. The other possibility, soft decision combining, is to merge the raw sensing samples and decide on the presence of a PU based on the merged raw samples. Between these two concepts, there also exist concepts where the sensing samples are preprocessed (e.g., one-bit decisions together with a confidence level) and the decision is based on the preprocessed samples. An overview of possible fusion techniques for sensing samples can be found in [26].

The three possible diversity dimensions discussed in this thesis are time (several sequential sensing samples taken by a single sensor in the same frequency band), frequency (several sensing samples taken by a single sensor at the same time in different frequency bands), and space (several sensing samples taken by different sensors (placed in different locations) at the same time in the same frequency band). Obviously, the different dimensions of diversity can be combined. Note that there is a strict limit on the achievable time and frequency diversity. In the time domain, the achievable diversity is limited by the maximum interference time (t_{max}) and t_{coh} , which limit the maximum number of i.i.d. sensing samples. In the frequency domain, the limits are the PU bandwidth (b_{pu}) and b_{coh} .

Although there are no such theoretical limits on the maximum achievable diversity in the spatial domain, there are certain practical things to be considered. For spatial diversity, often also referred to as *distributed sensing*, there are additional processes required before the sensing samples can be merged. First, it has to be decided who participates in the distributed sensing process. While this can be all CRs within the SUG, it might also be only a subset. The selected CRs then have to perform the local sensing process, gathering the sensing samples. These sensing samples then have to be exchanged to make a decision based on all sensing samples.

Conceptually, two different approaches can be followed within distributed sensing: a centralized and a distributed approach. In the centralized approach, a central controller decides on the subset of CRs to perform the sensing and all sensing samples are gathered at the central controller, which merges the samples, makes a decision based on the merged samples, and distributes the decision back to the other CRs. In the distributed approach, the sensing samples are exchanged between the CRs, and each CR makes its own decision based on its own and the received sensing samples.

2.3.2 Spectral Resource Management

In traditional networks (e.g., cellular GSM networks) spectral resource management is — so far — usually a static, offline task of, e.g., computing an optimal frequency assignment to minimize interference. CR networks, in contrast, require a highly dynamic spectral resource management. The most important requirement of spectral resource management in CR networks is the protection of PUs. SULs have to be reconfigured and adapted to the changing activity of PUs.

The highly dynamic spectral resource management required in CR networks also has great potential. Since the used resources have to be constantly adapted to the PU activity anyway, it becomes fairly natural to also adapt the used resources to the required throughput, leading to communication links which can dynamically grow and shrink based on the bandwidth requirements of the applications.

The above discussed dynamics create, however, an additional challenge: to maintain a certain QoS for a SUL. In traditional networks, once a link is set up, the available spectral resources usually do not change. This is not the case for CR networks. Used spectral resources have to be abandoned once a PU appears. Depending on the reconfiguration mechanisms it might take time until new resources are acquired, leading to (temporarily) fewer spectral resources available for the SUL, which in turn might severely degrade the secondary QoS.

In the following, first the construction and maintenance of Secondary User Links (SULs), being the spectral resources to be managed in case of CR networks, is introduced. Further, the two main mechanisms for spectral resource management within CR networks are explained in detail: link reconfiguration to adapt to the changing activity of PUs and secondary QoS maintenance mechanisms to compensate for spectral resources lost due to PU appearance.

Secondary User Links

A SUG, as introduced already above, creates a Secondary User Link (SUL), which is used for communication among all members of the SUG. A SUL is constructed out of spectral resources temporarily not used by the respective PUs. It can be entirely placed within one PU band, but can also use spectral resources from multiple PU bands. There are two possibilities for constructing SULs: contiguous SULs and non-contiguous SULs as shown in Figure 2.1.

A contiguous SUL consists of a contiguous amount of spectrum. Figure 2.1 shows three examples for contiguous SULs: one which occupies only part of a PU band (SUL 1), one which occupies a whole PU band (SUL 2), and one which occupies more than one PU band (SUL 3). The use of contiguous SULs is envisioned, e.g., within the IEEE 802.22 standardization [73].

Non-contiguous SULs consist of multiple, non-adjacent spectrum bands. One approach, introduced in [17], is to consider the PU bands as a bundle of smaller sub-channels. The idea behind the non-contiguous SULs is to scatter the SUL over multiple PU bands, such that (i) a reappearing PU is less affected (only a few sub-channels are

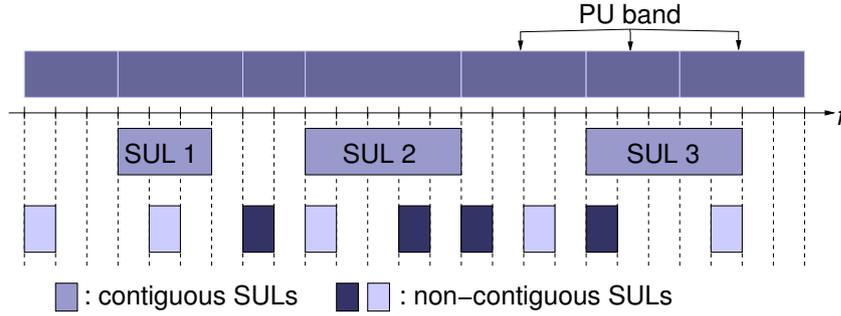


Figure 2.1: Examples for contiguous and non-contiguous Secondary User Links

used), and (ii) only a very small number of sub-channels has to be exchanged in the SUL if a PU appears. In Figure 2.1 two non-contiguous SULs are shown. For the rest of this thesis, we use the term “sub-channel” to refer to a chunk of spectrum used for a SUL.

The link maintenance process is responsible for the proper operation of a SUL during the whole duration of its usage. Generally speaking, link maintenance is the process of surveying the availability of the sub-channels used for the SUL and adjusting them in case a PU was detected. Providing a proper QoS for the SUL despite the necessary reconfigurations is an essential part of the link maintenance process. Conceptually, link maintenance can be divided into link reconfiguration, responsible for the release and adding of sub-channels to the SUL, and secondary QoS maintenance, responsible to compensate for potential temporal performance degradations of the secondary QoS due to reconfigurations.

Link Reconfiguration

Link reconfiguration is the part of the link maintenance process responsible for the adjusting of the sub-channels used for the SUL in case a PU was detected in one of these sub-channels.

The first part of link reconfiguration is to ensure that all members of a SUG stop transmitting on the sub-channels on which a PU was detected. In order to satisfy the protection requirements for the PU, it is crucial that transmission on the affected sub-channels is terminated within a certain time which does not violate the t_{\max} constraint. Reliability of the signaling is very important since even a single CR not receiving the PU detection notification can cause harmful interference.

In addition to stopping transmission on sub-channels on which a PU was detected, new sub-channels might be acquired to compensate for the lost ones. The decision of which new sub-channels to use is based on the sensing results indicating which sub-channels are available. Note that this might require to trigger a new sensing process. In order to add new sub-channels to the SUL (replacing possibly lost ones) it has to be ensured that these sub-channels are not used by any PU. Depending on the capabilities of the sensing process, this can be done proactively during the periodic sensing process, i.e., in parallel to sensing the sub-channels used by the SUL, or reactively upon request

of the link reconfiguration process. Whether sensing on potential new sub-channels is done proactively or reactively obviously has a big impact on the time required for link reconfiguration (t_{reconf}).

Whereas the time for vacation of resources is limited by the interference constraints, there is — in general — no such limit for the negotiation of new resources. However, note that during the link reconfiguration time (t_{reconf}) the spectral resources available for the SUL are reduced, influencing the secondary QoS. Although no harm is done to PUs if some CRs of the SUG do not receive the (correct) information about the new sub-channels to use, the affected CRs lose the connectivity. Re-establishing the connectivity might take a significant amount of time and, thus, has a severe impact on the quality of the secondary communication.

Link reconfiguration can be realized in a centrally controlled or distributed fashion, which has an influence on the complexity of the protocols, the required communication overhead, and t_{reconf} . The size of the SUG additionally has an influence on these parameters.

Secondary QoS Maintenance

QoS metrics used within the spectrum domain are usually related to the offered spectral capacity, i.e., the amount of available bandwidth within a certain time period. The offered spectral capacity can, e.g., be quantified by its average, its minimum, or by α quantiles, specifying the minimum capacity offered $\alpha\%$ of the time. In CR networks there are two processes which have a big influence on the QoS of a SUL: periodic spectrum sensing and link reconfigurations.

All spectral resources used by a SUL have to be periodically sensed to check whether a PU appeared and the resources have to be vacated. As explained above, spectral resources being sensed cannot be used for data transmission at the same time. The commonly followed approach of interrupting the communication on the SUL, perform spectrum sensing, and then resume the communication again can obviously severely degrade the secondary QoS. In order to maintain the secondary QoS also during the periodic sensing process, more advanced spectrum sensing approaches are required.

If a PU was detected, the SUL has to be reconfigured. Link reconfiguration requires time during which the spectral resources of the SUL are potentially reduced. Such bandwidth reduction can severely degrade the secondary QoS. To avoid such degradation of the secondary QoS, the link maintenance process has to ensure that even in case of reconfigurations of the SUL, the minimum amount of spectral resources is available with a very high probability. The main challenge for a SUG to maintain a certain secondary QoS is to cope with the highly variable availability of spectral resources.

2.4 Regulation, Standardization, and System Designs

In this section we give an overview of the past and ongoing regulation and standardization activities for opening up spectrum access for secondary usage, since those activities are

a driving force for the development and the deployment of CR systems. Furthermore, we give a brief overview of some of the most important CR system designs proposed in the research community.

2.4.1 Regulation Activities

Dynamic Spectrum Access does not only require the design and development of new, spectrum agile technologies. Since DSA fundamentally changes the traditional spectrum allocation, new regulatory policies and rules are also required. In 2003, the FCC has issued a Notice of Proposed Rule Making [46] advancing CR technology as a candidate to implement negotiated or opportunistic spectrum sharing. One year later, the European Commission assigned a “Study on Conditions and Options in Introducing Secondary Trading of Radio Spectrum in the European Community” whose results can be found in [7]. The focus with this study is on the question of to which degree secondary trading requires coordination across the member states of the European Union (EU).

The first regulation activity proposing a concrete regulating of secondary usage is the so called “white space” usage of the TV bands in the U.S. The FCC proposed the approval of secondary operation in the TV broadcasting bands in the U.S. [47] to provide broadband wireless access in rural areas. In this first proposed rule making the protecting of the TV broadcasters had to be ensured by a sensing process, augmented with detection based on a geo-location database of used TV channels. The sensing process was required to detect TV broadcasting at a level of -114 dBm. However, since there are also operating licensed and unlicensed wireless microphones within the TV bands, any CR system operating in the TV bands had to also ensure unimpaired operation of these microphones. Thus, for secondary operation in the TV bands, the FCC required the sensing process to also detect these wireless microphones.

Very recently, the FCC adopted the proposed rule making to not require spectrum sensing anymore [49]. Thus, secondary usage of TV bands follows the announced spectrum sharing approach. To still ensure the proper operation of wireless microphones, two TV channels are excluded from any secondary usage and reserved for wireless microphones (“safe harbor”).

In 2009, the European Conference of Postal and Telecommunications Administrations (CEPT) launched a new project (SE43) within the Spectrum Engineering Working Group (WG SE) to investigate the requirements for the operation of CR systems in the white spaces of the TV bands (470 – 790 MHz).

Regulation of secondary spectrum usage is still in its infancy. However, the activities summarized above show that regulatory bodies are starting to realize the need for a more flexible radio spectrum regulation.

2.4.2 Standardization Efforts

Standardization is an important driver for the detailed development and implementation of CR system designs. Without industry standards, the financial risks for companies to

invest in new approaches such as DSA is often too high. In the following, we summarize the most important standardization activities related to DSA and CR.

IEEE 802.22

The IEEE 802.22 working group develops standards specifically targeted at the FCC regulation for TV white spaces. It is the first standardization effort in the area of CR and secondary usage of spectrum. 802.22 systems are designed to offer broadband wireless access in rural areas utilizing unused TV channels. The general system architecture of 802.22 is based on the WiMAX standard (IEEE 802.16) [80]. As of now, the standard is still in draft status (draft version 6.1 beginning of 2011) [82].

In 802.22 a fixed base station offers wireless broadband to so-called Customer Premise Equipments (CPEs), which are typically also fixed. The typical transmission range of such a system is envisioned to be in the order of 30 km. The base station selects available channels consulting a geo-location database as required by FCC regulation. Note that, although spectrum sensing is not required anymore for operation in the TV bands, the 802.22 standard defines spectrum sensing processes and protocols to signal the sensing results from the CPEs to the base station. In fact, in 2010, the amendment IEEE 802.22.1 was published [81], which enhances the “Harmful Interference Protection for Low-Power Licensed Devices Operating in TV Broadcast Bands”.

SCC41

The IEEE Standards Coordinating Committee 41 (SCC41) is the group within IEEE responsible for the development of a series of 1900 standards for “Dynamic Spectrum Access Networks”. There are several working groups within SCC41, some of which already have published standards or have standards in a draft status.

The “IEEE 1900.1 Working Group on Definitions and Concepts for Dynamic Spectrum Access” is responsible for “Terminology Relating to Emerging Wireless Networks, System Functionality, and Spectrum Management” and has already released a standard [74]. The IEEE 1900.2 working group has released a standard on the “Recommended Practice for the Analysis of In-Band and Adjacent Band Interference and Co-existence Between Radio Systems” [75].

The IEEE 1900.4 working group is defining standards for the general architecture of CRs. A first standard defining the “Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks” was published in 2009 [76]. The working group is now working on an amendment for TV white spaces: “Architecture and Interfaces for Dynamic Spectrum Access Networks in White Space Frequency Bands”. Furthermore, they are investigating the “Interfaces and Protocols Enabling Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Networks”.

In the IEEE 1900.5 working group a “Policy Language and Policy Architectures for Managing Cognitive Radio for Dynamic Spectrum Access Applications” are developed, which is required to enable the policy driven operation of CRs.

Finally, the IEEE 1900.6 working group just published the draft standard on “Spectrum Sensing Interfaces and Data Structures for Dynamic Spectrum Access and other Advanced Radio Communication Systems”.

The above efforts show that there is significant interest in the standardization of CR approaches. This shows the massive interest of industrial and commercial entities which are usually the main players in standardization. The scope of most of the above standards (terminology — 1900.1, recommended practice for analysis — 1900.2, architectural building blocks — 1900.4) show that, in contrast to other wireless standardization, there is still a need to define and standardize fundamental concepts and architectures in the context of CR. However, it can also be seen that standardization also slowly moves towards more concrete systems (e.g., secondary usage in TV white spaces).

IEEE 802.11

Also within the IEEE 802.11 standardization there are efforts towards cognitively enabling the popular WLAN standard. In fact, the 802.11h amendment published in 2003 and meanwhile incorporated in the main standard [78] can be seen as the first step towards cognitively enabling 802.11. The amendment defines rules for dynamic frequency selection (DFS) and transmit power control (TPC) for the 802.11a Medium Access Control (MAC) operating in the 5 GHz ISM band. 802.11h enabled the use of 802.11a in Europe, which was previously not possible due to potential harmful interference to radar and satellite systems working in the same band.

The amendment 802.11k for “Radio Resource Measurement of Wireless Lans” [79] is strictly speaking not a standard for cognitive enhancement of 802.11. However, it defines rules and protocols to exchange measured information (such as load, list of neighboring Access Points (APs), or Received Signal Strength Indication (RSSI) measurements) between APs and between AP and stations. It is, thus, an important standard facilitating the development of cognitive enhancements for 802.11.

The 802.11y amendment [77] enables WLAN to operate in an announced spectrum sharing mode in the 3650 – 3700 MHz band in the U.S. In this modus, base stations are required to get a license, however, terminals do not need one. A terminal can only transmit after having received an enabling signal from the base station.

Beginning of 2010, the IEEE 802.11 working group created the 802.11af task group to enable 802.11 in the TV bands. These activities show that cognitively enhanced wireless communication to increase the efficiency in spectrum usage is also an important topic for existing and widely used wireless standards.

2.4.3 Cognitive Radio System Designs

In this section we give a brief overview of the most important CR system designs proposed since the invention of the CR concept. The CR designs are presented in chronological order of their first publication. A more detailed overview can be found in [4]. We further provide a comparison of the presented designs based on selected criteria. Those criteria are:

- Dynamic Spectrum Access (DSA) approach,
- PU protection approach,
- required changes to PU systems,
- secondary spectrum usage approach,
- special features.

(Over)DRiVE

The DRiVE (Dynamic Radio for IP Services in Vehicular Environments) [95, 175] and OverDRiVE (Spectrum Efficient Uni- and Multicast Over Dynamic Radio Networks in Vehicular Environments) [62, 96] projects define one of the very early CR system designs. In fact, strictly speaking, the objective is not the coexistence of PUs and SUs but rather the dynamic spectrum coordination among primary systems.

The main idea is that different Radio Access Networks (RANs) only get allocated the amount of spectrum currently required, which might vary over time and space. DRiVE is, thus, following the negotiated spectrum sharing approach. Whereas the focus within the DRiVE project was on contiguous allocation of spectrum, in OverDRiVE the possibility of non-contiguous spectrum allocation was investigated.

Table 2.2: (Over)DRiVE

DSA approach	negotiated spectrum sharing
PU protection	n/a
PU changes	PU systems need to be enhanced to, e.g., support dynamic spectrum allocation
secondary usage	support of non-contiguous allocation (OverDRiVE)
special features	n/a

DARPA XG

The U.S. Defense Advanced Research Projects Agency (DARPA) follows a very general approach and tries to define an architectural framework for the deployment of CR within the next generation (XG) program [34]. Their vision is to solve the problem of dynamic spectrum management in its generality by developing a CR framework [36]. In fact, the general approach taken is similar to that of the 1900.4 standardization group described above.

The goal is to have a regulatory approved framework to avoid the often time-consuming system by system regulation and enable the fast deployment of new CR technologies. The XG framework is policy-centric, i.e., an XG system can be entirely controlled and predicted with policies and thus policies are the only thing regulatory

bodies need to concern themselves with. This also allows radios to transcend regulatory borders with simple policy changes.

In order to achieve this goal, DARPA-XG defines a core set of abstract behaviors, the “regulatable kernel”. The regulatable kernel only contains behaviors that are subject to regulatory approval, which are mainly behaviors ensuring that the system avoids interference with licensed systems.

In [35], the architectural framework is described. The three main modules within an XG node are opportunity awareness, opportunity allocation and opportunity use. Opportunity awareness is the only module subject to regulatory approval and can be decomposed into sensing, interpretation and dissemination of interference information. The opportunity allocation module then allocates the available resources determined by the opportunity awareness module. Finally, the opportunity use module is responsible for sending the bits using the allocated resources.

Table 2.3: DARPA XG

DSA approach	opportunistic spectrum sharing
PU protection	spectrum sensing based
PU changes	none
secondary usage	n/a
special features	clustering of regulation relevant functionalities into “regulatable kernel”

Spectrum Pooling

Spectrum pooling, proposed in [155], is an opportunistic spectrum sharing approach and assumes usage of non-contiguous SULs. The authors define a centralized spectrum pooling architecture based on Orthogonal Frequency Division Multiplexing (OFDM) assuming an Access Point (AP)-based WLAN scenario for the CR system. Time Division Multiple Access (TDMA) is used for multiple access within the SUG.

Each CR performs local spectrum sensing and the sensing results are gathered at the AP. For the timely and efficient signaling of the results to the base station a so called boosting approach has been proposed [157]. The main idea is to signal the sensing data on the physical layer by just emitting additional energy on sub-channels where PUs are present. The AP then gets an amplified signal and, based on that, can decide on the state of the sub-channel.

The advantage of the distributed detection over only the AP performing the sensing is discussed in [156]. The performance of the proposed system dependent on the usage patterns of the licensed systems is discussed in [22].

CORVUS

CORVUS [17, 20] is an opportunistic spectrum sharing approach developed throughout the course of this thesis in cooperation with the University of California, Berkeley. The

Table 2.4: Spectrum Pooling

DSA approach	opportunistic spectrum sharing
PU protection	spectrum sensing based
PU changes	none
secondary usage	OFDM-based, support of non-contiguous allocation
special features	physical layer sensing information exchange protocol

author of this thesis was one of the main developers of CORVUS.

Within CORVUS non-contiguous SULs are used for the communication within SUGs as explained in Section 2.3. Sensing is assumed to be done by the CRs, so no external sensing infrastructure is required. In CORVUS, *logical* control channels are defined, which are used for the exchange of control information within a SUG and between different SUGs. Ultra-Wideband (UWB) has been identified as a potential candidate for the physical implementation of these control channels. The details of the CORVUS system design relevant for this thesis are described in Chapter 6.

Table 2.5: CORVUS

DSA approach	opportunistic spectrum sharing
PU protection	spectrum sensing based
PU changes	none
secondary usage	OFDM-based, support of non-contiguous allocation
special features	UWB-based control channels

Spectrum Agile Radio

The spectrum agile radio system design [103] is based on the DARPA XG concept introduced above. Following the XG approach, the focus is on policies and a machine understandable language to represent these policies such that a CR can autonomously apply them. Spectrum agile radios download policies from servers or memory devices. These policies define the general operation rules for the radios. The policies describe the rules regarding which spectrum is available for secondary usage and under which conditions. Based on these rules, the CRs then autonomously manage their spectrum usage. CRs also disseminate the policies such that other CRs can learn about the policies.

Whereas the general idea of the specification, distribution, and application of policies is discussed in great detail, no concrete usage scenarios are discussed. The general approach can be applied to opportunistic or announced spectrum sharing and also many different secondary usage approaches are possible.

In [104] the spectrum agile radio idea is extended to include coexistence awareness among different SUGs. The coexistence approach proposed for spectrum agile radio is based on social awareness and game theoretic approaches.

Table 2.6: Spectrum Agile Radio

DSA approach	announced or opportunistic spectrum sharing
PU protection	not specified (but sensing or announcement-based possible)
PU changes	potentially none
secondary usage	n/a
special features	game theory-based coexistence between CRs

DIMSUMnet

The dynamic intelligent management of spectrum for ubiquitous mobile network (DIMSUMnet) [18, 19] is a system approach for negotiated spectrum sharing. Similar to the OverDRiVE approach, DIMSUMnet does not differentiate between PUs and SUs but rather focuses on increasing the efficiency of spectrum usage by novel spectrum allocation approaches.

DIMSUMnet assumes a coordinated access band (CAB) of continuous spectrum which is owned by a so called “spectrum broker”. The spectrum broker auctions the spectrum to the network operators or individual users. One of the advantages of such an approach over each operator having its own spectrum band is a more efficient spectrum usage through the statistical multiplexing gain.

The DIMSUMnet architecture consists of a spectrum information and management (SPIM) broker, RAN managers (RANMAN) and DIMSUMnet enabled base station and end user devices. When a base station boots, it contacts its RANMAN, which negotiates a lease with the SPIM broker and configures the base station. The base station can then offer its services to the end user devices. Two control protocols are proposed: spectrum Lease (SPEL) protocol for communication between base station, RANMAN and SPIM broker and Spectrum Information Channel (SPIC) protocol between the base station and the end user devices.

Table 2.7: DIMSUMnet

DSA approach	negotiated spectrum sharing
PU protection	n/a
PU changes	broker and RAN manager required; base station need to be enhanced to be able to talk to broker and manager
secondary usage	auction-based
special features	n/a

IEEE 802.22

The IEEE 802.22 standard follows the announced spectrum sharing approach. It standardizes the secondary usage of white spaces in the TV bands (see Section 2.4.2 for details). Currently, only contiguous allocation (of up to one complete TV channel) is supported; however, the “bundling” of multiple TV channels is also been discussed.

Table 2.8: IEEE 802.22

DSA approach	announced spectrum sharing
PU protection	geo-location database
PU changes	server for geo-location database required
secondary usage	only contiguous allocation, based on 802.16 (WiMAX)
special features	n/a

2.5 Definitions

In this section we summarize the main definitions used throughout the thesis and provide a pointer to the context in which they are defined.

Primary User (PU): A Primary User is the license holder of some spectrum. It has strict priority in access to this spectrum. (Defined in Section 1.1.)

PU band: A PU band is the spectrum band owned and used by a PU for its communication. (Defined in Section 1.1.)

Cognitive Radio (CR): A Cognitive Radio is a radio that can change its transmitter parameters based on interaction with the environment in which it operates. The majority of CRs will probably be Software Defined Radios (SDRs), but neither having software nor being field programmable are requirements of a CR. (Based on the FCC definition [46] and defined in the context of Section 2.1 in this thesis.)

Secondary User (SU): A Secondary User can use spectrum of a PU band only if the respective PU is not present, with the constraint that it has to vacate the spectrum again as soon as the PU returns. (Defined in Section 1.1.)

Secondary User Group (SUG): A Secondary User Group is defined as a set of CRs (terminals or a base station with associated terminals) which form a group and jointly use temporarily unused spectrum for their own communication purposes. Members of a SUG usually also cooperate in the sensing process. (Defined in Section 2.3.)

Dynamic Spectrum Access (DSA): Within this thesis, Dynamic Spectrum Access defines approaches used by SUGs to fill the gaps created by the static spectrum access of the PUs and utilize the temporarily unused spectrum. (Defined in Section 2.2.)

Opportunistic Spectrum Sharing: Opportunistic spectrum sharing is a DSA approach, in which spectrum access is based on identifying the presence of PUs by a sensing process. The rules and conditions for this sensing process as well as the usage of the spectrum are defined by (regulatory) policies. The sensing is mostly done by the CRs themselves; however, sensing can also be done (or complemented) by external sensing devices specifically deployed by third parties (e.g. regulators or the PU itself). (Defined in Section 2.2.3.)

sub-channel: A sub-channel is a chunk of spectrum used for a SUL. It is the smallest allocation unit that can be used. (Defined in Section 2.3.2.)

Secondary User Link (SUL): Spectrum used by a SUG is referred to as a Secondary User Link and assumed to be used in a shared modus to support communication among all the CRs involved. A SUL can consist of one or multiple sub-channels. (Defined in Section 2.3.)

maximum interference time (t_{\max}): The maximum interference time is defined as the maximum amount of time for which a certain PU, resuming its activity after a break in spectrum usage, can tolerate interference from secondary usage. The CR system has to ensure that the PU band is vacated within t_{\max} after a PU appears in its PU band. (Defined in Section 2.2.3.)

sensing sample: A sensing sample is obtained by sampling the received radio signal with a frequency span of b_{coh} and for a time of t_{coh} . Note that the sampling frequency used to sample the radio signal is usually different from the frequency with which sensing samples are recorded. The received signal is usually sampled at Nyquist rate and many of these micro-samples are used to obtain one *sensing sample*. (Defined in Section 2.3.1.)

sensing reliability: Sensing reliability is a measure for the interference created by a SUG to the PU network. It is quantified by the probability of false negatives (P_{fn}), i.e., the probability of not detecting a PU in a pre-defined time period (t_{\max}) although it is present. To ensure the protection of the PU, the sensing process has to ensure to keep P_{fn} below a certain threshold (usually defined by regulation). (Defined in Section 2.3.1.)

sensing quality: Sensing quality is a measure for the effectiveness of the sensing process in detecting idle spectral resources. It is quantified by the probability of false positives (P_{fp}), i.e., the probability of detecting a PU although it is not present. (Defined in Section 2.3.1.)

CHAPTER 3

**Thesis Motivation and
Approach**

3.1 Thesis Motivation

Cognitive Radio, being a relatively young research area, still faces a lot of challenges and has a lot of open issues even on the general system design side. However, there is already a lot of research going on and it is to be expected that the first systems, operating in the TV white spaces to provide broadband wireless access, will be rolled out soon.

Those first systems, although expected to provide lots of insights for future research, have certain limitations. The targeted spectrum bands are owned by very static PUs. Spectrum occupancy on TV channels is likely to be stable for long time periods and, generally, shows a very predictive behavior. Furthermore, the most recent FCC ruling [49] requires the PU detection only to be based on geo-location database-assisted detection and eliminates the need for spectrum sensing. CR networks in these bands will be rather static, have geo-location information available, and most likely not have strict constraints on processing capabilities or energy constraints.

In this thesis we address the problem of secondary spectrum usage in more dynamic spectrum environments focusing on resource (complexity, processing, power) limited CR systems. Admittedly, it could be argued that due to the abundance of available spectrum in PU bands with rather static spectrum availability (as, e.g., TV bands), secondary usage in PU bands with a very high variability of available spectrum has no valid use case. However, we strongly believe in the relevance of secondary usage also in those PU bands mainly due to three reasons.

Firstly, we believe that the secondary usage scenario is only an intermediate use case for CR. Ultimately, even if there is still a distinction between primary and secondary networks, both will use cognitively enabled technologies trying to maximize the spectrum utilization and to minimize the impact on other networks. We, thus, believe that concepts are required which enable CR systems to also operate in very dynamically changing spectrum environments. For such scenarios, PU detection based on occupancy databases is not really an option. The focus of this thesis is, thus, on opportunistic spectrum sharing based on spectrum sensing.

Secondly, the heavily used PU bands showing a high variability in available spectrum (e.g. cellular bands) are used worldwide, resulting in a large number of deployed networks and devices. Consequently, the hardware and technologies used are well understood and available at reasonably low costs. Note that this is also attractive for wireless operators to extract additional revenue from their licensed spectrum. Following this approach, CR systems do not have to be designed and developed from scratch, reducing costs and time-to-market for initial CR deployments. Furthermore, there has been recently significant interest in implementing DSA in such bands (e.g. [5, 6, 18, 24, 66, 170]).

Thirdly, as we will show in Chapter 4, the subjective perception of the variability of available spectrum by the CR system can be significantly higher than the variability due to PU agility. On the one hand, false positives of the sensing process result in an increased variability of spectrum as perceived by the CR system and, on the other hand, also mobility of the primary and / or secondary networks increases the perceived variability of available spectrum. We believe that CR approaches are especially attractive for small

WLAN-like networks to satisfy the ever increasing demand for wireless connectivity at “the last hop”. Many of those networks will be mobile resulting in an increased variability of available spectrum perceived by the CR system.

Another concept which we believe to be essential for the rapid deployment and evolution of CR networks is the possibility to gradually increase the performance of deployed CR networks. To make secondary spectrum usage a reality, the experience gained from real, operating networks is needed. The ability to initially deploy small scale and cheap CR networks is crucial to gain such experience. Obviously, even those initial deployments need to adhere to the regulatory constraints; however, the performance experienced in the secondary networks probably will not be optimal. To not waste spectral capacity as the number and size of CR networks increases, CR networks should be able to evolve, i.e., be able to increase their performance without the need to completely replace the old CR systems.

Summarizing, the three main questions this thesis answers are:

1. **Is the use case for secondary usage in PU bands with high variability of available spectrum justified?**
2. **How can sensing-based secondary spectrum usage be realized in scenarios where SUGs have to deal with high variability of available spectrum?**
3. **How should we design systems that enable secondary usage with simple, low cost devices without blocking the possibility to gradually enhance the system and its performance?**

3.2 Thesis Approach

We first justify our assumption of scenarios with high **variability of available spectrum** by investigating and comparing in detail the real and the measured spectrum usage by PUs. It is shown that, even for relatively static PUs, there can be significant variation in spectrum availability perceived by the CR network.

Based on this observation and to evaluate the performance of candidate systems for such challenging secondary usage scenarios, a triangle of three fundamental performance metrics is formulated.

The most important enabling functionality for secondary spectrum usage is the **reliable protection of the PU communication**. It has to be ensured that SUGs only utilize PU bands which are currently not used by any PU. Furthermore, the quick and reliable vacation of PU bands reclaimed by a PU has to be ensured.

Recognizing the requirement of reliable PU protection, DSA and secondary usage of spectrum can only become a service reality if at least some degree of **QoS support for the secondary communication** is provided, which is the second performance metric. Due to the requirement to always avoid PUs, secondary communication potentially often has to be relocated to different, currently available, spectrum. Thus, the fundamental

secondary QoS metric is to which extent service interruptions can be avoided and / or compensated for.

The third performance metric to be considered is the **spectral efficiency of the secondary usage**. The driving force for DSA and secondary usage of spectrum is the inefficient spectrum utilization of legacy systems. Whereas it could be argued that any additional usage improves the spectral utilization, innovative DSA approaches should be designed with the goal of efficient spectrum usage in mind. In particular, spectral efficiency becomes critical as soon as the demand for spectrum is bigger than its availability.

Having defined the metrics above, it is important to understand that reliable PU protection is a hard constraint for the deployment of CR systems and is, thus, the enabling metric for secondary spectrum usage. No secondary usage can take place if the PU protection requirement is violated. The other two metrics are soft constraints, serving as performance metrics for CR systems meeting the PU protection requirement.

3.2.1 Variability of Available Spectrum

We provide a detailed analysis of recent spectrum measurement campaigns, trying to extract information on the variability of available spectrum. Unfortunately, there is only very limited information about the variability of available spectrum in the investigated measurement campaigns. Even worse, those measurement campaigns only provide *estimates* of the available spectrum (and its variability) which are not validated against the real usage by the PUs.

We, thus, conducted a large scale measurement campaign, determining the real spectrum availability as measured *inside* a primary network. Within this campaign, we also collected detailed information about the variability of the available spectrum, which is used to develop a detailed PU spectrum usage model. The developed model is used to simulate the primary usage in Chapter 7 of this thesis.

Being able to measure the real spectrum usage, the question arises of if and how the real spectrum usage can be estimated through spectrum measurements. In contrast to the evaluated measurement campaigns, we conducted a parallel measurement of the real spectrum usage as measured inside the primary network and of the estimated spectrum usage as measured by spectrum sensors. The results of this measurement campaign show that in order to not overestimate availability of the unused spectrum, i.e., ensuring a low probability of false negatives (P_{fn}), potentially a high probability of false positives (P_{fp}) has to be accepted.

From the secondary usage point of view, a false positive is simply an increase of the variability of available spectrum. A CR cannot differentiate between a PU which appeared in its PU band and the sensing process reporting a false positive. Having a high probability of false positives is perceived by the CR as a high variability of available spectrum. We, thus, believe that for realistic secondary usage scenarios not only the dynamics of the PU usage have to be taken into account to determine the variability of available spectrum, but also the variability introduced by an (imperfect) sensing process responsible for the detection of the PUs.

Mobility (of the PUs and / or the CRs) makes things even worse. The higher the relative speed between both systems, the faster sensing data gathered by the CRs is outdated (e.g., a detected PU could already be out of interference range again). Following this line of argument, high variability of available spectrum is a real and valid use case for secondary spectrum usage.

3.2.2 Reliable Primary User Protection

Motivated by the results of the experiments estimating available spectrum, we investigate — from an information theoretic point of view — if and how reliable detection of PUs is possible with simple, energy detection based sensors. The theoretic results show that reliable detection, quantified by the probability of false negatives (P_{fn}), is also possible maintaining a good quality of the sensing process, quantified by the probability of false positives (P_{fp}). Those results are in contrast to our initial experimental investigations, which showed a very high P_{fp} . One of the reasons for the very high P_{fp} in our experimental investigation is that it was only based on the measurements from one single spectrum sensor. However, to have both a low P_{fn} and a low P_{fp} , a lot of *diversity* in the spectrum measurements is required. Diversity of spectrum measurements is achieved by combining multiple, independent measurements in order to make the decision on the presence (or absence) of a PU more robust to attenuation caused by the wireless channel. Three possible diversity dimensions are taking multiple independent measurements in the time domain, in the frequency domain, or in the spatial domain (from multiple spatially diverse sensors). In fact, time and frequency diversity alone are not sufficient to achieve a low P_{fn} **and** P_{fp} , i.e., the spatial diversity component is crucial. Limiting the (spatial) diversity in the spectrum measurements unavoidably results in an increased probability of false positives.

The case of very limited (spatial) diversity in the spectrum measurements is of high practical relevance since we want to be able to achieve reliable PU protection also for the case of initial CR deployments, only consisting of very few sensors (limiting the spatial diversity). Furthermore, agile PUs should also be reliably detected (limiting the time diversity). From the theoretical results we show that even in the case of very limited diversity, reliable PU protection can be achieved. The price for assuring a low probability of false negatives with only very limited spatial diversity is to accept a very high probability of false positives.

Based on the result that reliable PU detection is possible, we investigate the influence of diversity in the time, frequency, and spatial domain on the quality of the sensing process. We analytically and experimentally quantify the achievable diversity in the different domains and show tradeoffs between the achievable diversities for selected scenarios.

3.2.3 Secondary Quality of Service Support

Although secondary QoS support is not a hard constraint, a CR system should be able to provide some degree of QoS support to be successful. Secondary QoS support is a

big challenge for CR systems operating in PU bands with high variability of available spectrum, since the Secondary User Link (SUL) has to be continuously adapted to the varying spectrum availability. There are two major challenges secondary QoS support has to deal with. First of all the PU has to be detected by the sensing process, which requires interruption of the secondary communication in the spectrum to be sensed. Second, after the detection of a PU, the SUL has to be reconfigured, i.e., the sub-channels located in PU bands where a PU appeared have to be replaced. In the case of a high variability of available spectrum, these reconfigurations of the SUL have to be done very often. Remember that not only the agility of the PU contributes to the variability of available spectrum, but also the mobility of the PUs and CRs and the false positives of the sensing process. False positives also require a reconfiguration of the SUL, since the sensing process only reports the detection of a PU and can not differentiate between true positives and false positives. As seen in the spectrum sensing investigation, the probability of false positives can be really high in case of agile PUs, which significantly increases the necessary reconfigurations and, thus, the challenges for secondary QoS support.

This thesis presents a CR system design which addresses the challenges above and is able to provide a certain level of secondary QoS support even in environments with very high variability of available spectrum. To quantify the secondary QoS support, the minimum spectral capacity offered by a SUL is used as a QoS metric in the presented thesis. The proposed system design features a periodic sensing approach, which greatly reduces the impact of the sensing process on the secondary QoS. Furthermore, it uses a reconfiguration mechanism, which minimizes the impact of the appearance of individual PUs on the secondary QoS. In addition, approaches to compensate capacity losses of the SUL due to reconfigurations are proposed.

The impact of PU agility and false positives of the sensing process on the secondary QoS is investigated for the proposed system design. The investigations are conducted for different classes of PUs, representing different levels of agility. The PU occupancy models are based on the investigation presented in Appendix A. Our investigations show that the proposed system design can operate in PU bands with a wide range of occupancy dynamics and still provide a reliable secondary QoS support.

3.2.4 Spectral Efficiency

Both reliable PU protection as well as secondary QoS support require spectral overhead and have, thus, an influence on the spectral efficiency of the secondary usage. The probability of false positives (P_{fp}) of the sensing process also influences the spectral efficiency since idle PU bands wrongly declared as occupied cannot be used for secondary communication. P_{fp} further links the spectral overhead required for spectrum sensing and secondary QoS support and creates a tradeoff between the two. Using little spectral overhead for sensing results in a high P_{fp} which increases the spectral overhead required for secondary QoS support and vice versa.

Initial deployments of CR systems will most likely not be constrained by their spectral efficiency. Due to the abundance of temporarily unused spectrum, it is affordable to be

very spectrum inefficient in order to meet the other two metrics of reliable PU protection and secondary QoS support. However, as the density of CR system and the secondary traffic grows, spectral efficiency becomes an issue. Thus, even initial CR systems should be designed keeping spectral efficiency in mind.

The CR system design presented in this thesis is developed with the goal of supporting spectral efficiency, if required. Whereas small, initial network deployments focus on PU protection and secondary QoS support, the system is designed to be able to increase the spectral efficiency as the network grows and / or includes more sophisticated hardware.

We present an analysis of the spectral efficiency of both the spectrum sensing and the secondary QoS support approaches introduced above. The tradeoffs of where to spend spectral overhead is analyzed and quantified. Furthermore, it is shown that the introduced system design is scalable in the sense that it provides better spectral efficiency as the network size increases.

CHAPTER 4

Identification and Characterization of Available Spectrum

An overview of recent spectrum measurement campaigns in the context of CR is presented and the results are analyzed with respect to their usability to identify available spectrum and to characterize the dynamics of spectrum availability. To get an idea of the real spectrum usage compared to the above measurement campaigns, an investigation of the spectrum occupancy and dynamic behavior of cellular Primary Users (PUs), which is based on data gathered inside the PU network and published in [170, 171], is presented. Finally, an experimental investigation — published in [173] — is presented, which correlates spectrum measurements with the real usage recorded inside the PU network. It is shown that the variability of available spectrum not only depends on the PU agility, but also on the quality of the sensing process and the mobility of the primary transmitter and the CRs.

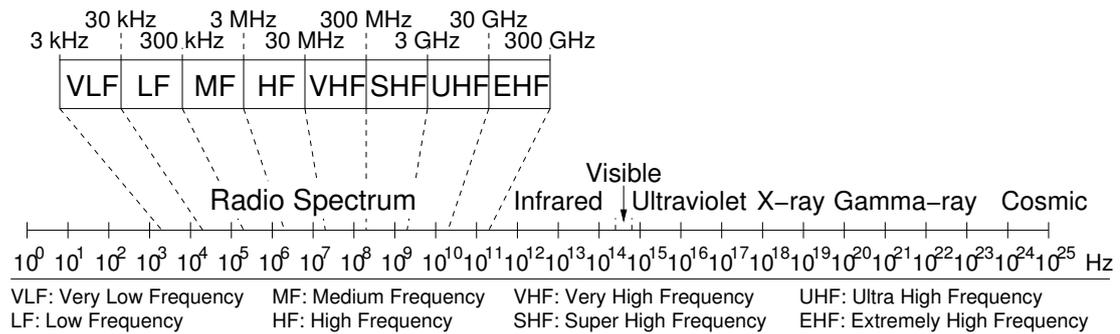


Figure 4.1: Electromagnetic spectrum

The part of the electromagnetic spectrum which can be used for wireless data transmissions is usually referred to as *radio spectrum* and defined as the frequency range of the electromagnetic spectrum where an alternating current fed to an antenna can generate electromagnetic waves (3 kHz to 300 GHz)¹. Figure 4.1 shows the electromagnetic spectrum from 0 to 10^{25} Hz and the location of the radio spectrum.

The International Telecommunication Union Radiocommunication Sector (ITU-R) subdivided the radio spectrum into eight frequency bands from Very Low Frequency (VLF) to Extremely High Frequency (EHF). Depending on their frequency, the bands have different propagation characteristics and are thus used for different purposes. Lower frequencies have a bigger propagation range compared to higher frequencies (assuming the same transmission power). The data-rate that can be supported on the other hand increases for higher frequencies, which is due to the shorter periodic time of radio waves at higher frequencies. A shorter periodic time means that the same amount of information can be transmitted in a smaller amount of time. Another reason is that there is significantly more spectrum available at higher frequencies. Comparing the available bandwidth of the Medium Frequency (MF) and the Very High Frequency (VHF) bands for example, shows that the amount of usable spectrum for the VHF band is 100 times the amount of the available spectrum for the MF band (note the logarithmic scale for the frequency in Figure 4.1). The shorter periodic time at high frequencies and the larger amount of spectrum at these frequencies justify the statement that — generally speaking — transmissions at higher frequencies can carry a bigger amount of data than transmissions at lower frequencies.

The by far most popular bands for wireless transmissions today are the VHF and Super High Frequency (SHF) bands. This is mainly due to the good propagation characteristics (transmission range, building penetration, etc.) of these frequencies. Note that there are also some technologies operating in the High Frequency (HF) band (e.g., AM radio broadcast), the Ultra High Frequency (UHF) band (IEEE 802.11a, wireless

¹Infrared transmission (frequently used in remote controls) is an exception here. Recently, also the visible light spectrum is considered for wireless transmissions in the research community. This new research trend is commonly referred to as “optical wireless”. The focus of this thesis, however, is on “traditional” wireless communication in the radio spectrum.

USB, planed UMTS / 4G extension bands) and the EHF band (WirelessHD). The main focus of today's CR research, however, lies in the VHF and SHF bands.

4.1 Spectrum Measurement Campaigns

Recent spectrum measurements in the 30 – 3000 MHz range are summarized in this section. Some of these measurement campaigns also include the spectrum up to 6 GHz; however, the focus in this thesis is on the spectrum up to 3 GHz. A special emphasis is put on the TV bands and the cellular bands since there are often more detailed results available for those bands.

4.1.1 Measurement Methodology

The vast majority of measurements to estimate the spectrum occupancy by Primary Users is done using general purpose, high precision spectrum analyzers. Usually different antennas are used for the different frequency ranges to be covered. In most of the presented measurement campaigns the measurement equipment is located on the rooftop of high buildings in urban areas. In some campaigns, the outdoor measurements are complemented by indoor measurements in office environments. In all presented campaigns the measurements are conducted at one single location at a time. Although there are some campaigns which analyze and compare different measurement locations, the measurements in these cases are done sequentially in time. An exception is the campaign presented in [24], where *concurrent* measurements are done in four different locations.

The spectrum analyzers record the Power Spectral Density (PSD) of the measured spectrum band over time. The recorded PSD samples are then split into smaller spectrum bands based on the class of allocated PUs (e.g., TV bands, cellular bands, etc.). Note that these smaller spectrum bands usually still consist of several PU bands, i.e., several PUs allocated in adjacent PU bands are jointly evaluated. The spectrum band from 470 to 512 MHz in the U.S., e.g., consists of 7 TV bands, each spanning 6 MHz.

In order to estimate which parts of the spectrum are occupied by PUs and which are unused, a detection threshold (τ) is used. If the measured PSD exceeds this threshold, the respective band is assumed to be occupied; if not, it is assumed to be unused. In the campaigns summarized below, τ is determined either based on noise measurements or based on the calculation of the theoretical ambient noise floor. Note that τ may vary for different spectrum bands. Let $\Phi_{i,t}$ be the Power Spectral Density of spectrum band i at time t , then the estimated binary spectrum occupancy variable of spectrum band i at time t ($\Psi_{i,t}$) is defined as

$$\Psi_{i,t} = \begin{cases} 0 & \text{if } \Phi_{i,t} < \tau_i , \\ 1 & \text{if } \Phi_{i,t} \geq \tau_i . \end{cases} \quad (4.1)$$

The estimated average duty cycle of spectrum band i (D_i) is a measure for the fraction of time an observed spectrum band i is occupied by PUs and is used as a metric for *spectrum occupancy* in most measurement campaigns. Let $t \in T$ be the observation

points within the observation period T . The estimated average duty cycle of spectrum band i can then be calculated as

$$D_i = \frac{\sum_{t=1}^T \Psi_{i,t}}{T}. \quad (4.2)$$

4.1.2 Wide-band Spectrum Measurements

The Chicago measurements [102, 109] are part of a measurement campaign in the U.S. to investigate the spectral occupancy of the radio spectrum by PUs. The measurements were conducted in 2005 during two days on a rooftop in downtown Chicago (Chicago Loop) and investigate the spectrum occupancy between 30 and 3000 MHz. The duty cycles of the investigated sub-bands are summarized in column 1 (“Chicago I”) of Tables 4.2 and 4.3. The Chicago measurements are part of a bigger campaign to perform spectrum measurements in multiple U.S. locations which started in 2004. Further results of this campaign for various locations in the U.S. can be found at the Shared Spectrum Company (SSC) website [137]. The SSC measurements are also the model for a number of other wide-band spectrum measurements which, in the following, are summarized in chronological order.

A large set of spectrum occupancy data was gathered in the Aachen measurement campaign in Aachen, Germany and Maastricht, Netherlands between December 2006 and July 2007 [158–161]. The campaign covers the frequency range from 20 to 6000 MHz, divided into four sub-bands and provides one week of measurement data for each sub-band in three different locations: the ground floor of an office building in Aachen, the third floor balcony of a residential building in Aachen, and the rooftop of the main building of the International School Maastricht in Maastricht. The measurement traces of this campaign are available online (registration necessary) [135]. Some first analysis of the indoor and outdoor measurements in Aachen of the above campaign are presented in [160]. However, a detailed analysis of the spectrum occupancy of the whole measured frequency range, as presented in the SCC measurement campaigns, is missing, so that we cannot provide this data.

Table 4.1 summarizes the results of a measurement campaign in urban Auckland, New Zealand [27] published in 2007. The measurement was done in the morning and afternoon within a 12-week period in the 806 to 2750 MHz frequency range. The measurement campaign comprises two sets of measurements, one taken on the rooftop on the ninth floor and the second one indoors on the third floor. Unfortunately, the exact frequency ranges of the investigated services are not specified.

As a result of the first Chicago measurement summarized above, there is now a permanent measurement setup in Chicago on the rooftop of the 22-story building of the Illinois Institute of Technology (IIT), located 5.3 km south of the Sears Tower. The setup is able to monitor the spectrum between 30 and 6000 MHz. Results from measurements taken between December 2007 and June 2008 are presented in [11]. The average occupation results are summarized in column 2 (“Chicago II”) of Tables 4.2 and 4.3.

A set of 24-hour measurements during 12 weekdays carried out on the rooftop of the

Table 4.1: Spectrum Occupancy in New Zealand (2007) in the 806 to 2750 MHz frequency range

Type of allocation	Outdoor	Indoor
Fixed linking service	6.67 %	4.49 %
Land mobile radio	8.95 %	4.41 %
uplink	3.6 %	3.33 %
downlink	19.04 %	9.1 %
Cellular mobile service	33.37 %	25.93 %
uplink	7.62 %	5.75 %
downlink	64.15 %	50.05 %
Aeronautical radio	2.35 %	8.47 %
Other	3.83 %	1.84 %

6-story Infocomm Research building in Singapore are presented in [85] published in 2008. The measurements characterize the spectrum occupancy between 80 and 5850 MHz and are summarized in the third column (“Singapore”) of Tables 4.2 and 4.3.

A measurement campaign conducted in the Guangdong province in China analyzing the spectrum occupancy *concurrently* measured in four locations over a duration of one week during February 2009 is presented in [24]. All four measurements were taken at the rooftops of buildings; two in downtown Guangzhou (roughly 10 km apart) and two in suburban areas (roughly 45 km apart). Although the frequency range from 20 to 3000 MHz is measured, only the cellular Code Division Multiple Access (CDMA), GSM, and TV bands are analyzed. This is why no overview results are presented here; however, details on the TV and cellular bands are summarized in the following sub-sections.

A wide-band measurement conducted in Europe is presented in [101], published in 2009. The authors have done a 48-hour measurement of the 75 to 3000 MHz spectrum in urban Barcelona, Spain. The measurements were taken on the rooftop of the north campus of the Universitat Politècnica de Catalunya (UPC). The results are shown in column 4 (“Spain”) of Tables 4.2 and 4.3.

Another European measurement campaign, was done in Bucharest, Romania, at the rooftop of the Telecommunications Department main building of the Politehnica University of Bucharest and published in 2010 [105]. The measurement campaign was done in the 25 to 3400 MHz frequency range and is summarized in the fifth column (“Bucharest”) of Tables 4.2 and 4.3.

The general trend of all the results is the same: based on the measurements, there is a significant amount of unused spectrum in all locations throughout the whole investigated spectrum range. Generally, the spectrum up to 1000 MHz tends to have more spectral activity than the spectrum between 1000 and 3000 MHz.

In the spectrum range up to 1000 MHz, broadcasting services such as FM radio and TV broadcasting are located. The usually constant emission of relatively high power explains the high estimated occupancy. In addition to broadcasting services, also the

Table 4.2: Spectrum Occupancy Table 30 – 960 MHz

	Chicago I (2005)	Chicago II (2007)	Singapore (2008)	Spain (2009)	Bucharest (2010)
	30-54 MHz 21.20% PLM, Amateur, others	30-54 MHz 22.00% PLM, Amateur, others			25-230 MHz 28.44% FM radio, Aero/Marine, Fixed/Mobile, Military, other
	54-88 MHz 70.90% TV 2 -6, RC	54-88 MHz 54.00% TV 2 -6, RC	80-174 MHz 35.00% FM Radio, Aero, Fixed/Mobile, others	75-174 MHz 90.00% FM radio, maritime, aero mobile/radionav, amateur, PMR, paging	
	108-138 MHz 2.60% Air traffic Control, Aero Nav	108-138 MHz 4.00% Air traffic Control, Aero Nav			
	138-174 MHz 35.20% Fixed Mobile, amateur, others	138-174 MHz 9.00% Fixed Mobile, amateur, others			
	174-216 MHz 44.80% TV 7-13	174-216 MHz 44.00% TV 7-13	174-230 MHz 49.00% TV, DAB	174-223 MHz 100.00% Audio broadcasting DAB-T, wireless microphones	
	216-225 MHz 4.40% Maritime Mobile, Amateur, others	216-225 MHz 7.00% Maritime Mobile, Amateur, others		223-235 MHz 62.00% Trunking	
	225-406 MHz 2.70% Fixed Mobile, Aero, others	225-406 MHz 0.00% Fixed Mobile, Aero, others	230-406 MHz 4.00% Fixed/Mobile, PMR, Aero, others	235-400 MHz 37.00% Ministry of Defense	230-400 MHz 11.09% Military, Mobile
	406-470 MHz 17.20% Amateur, Radio Geolocation, Fixed, Mobile, Radiolocation	406-420 MHz 5.00% government	406-490 MHz 16.00% Mobile Data, PMR & Trunk	400-470 MHz 38.00% Satellite, fixed and mobile (SRDs, paging, ISM-433, PMR-446)	400-470 MHz 18.37% Analogue/Digital, Terrestrial Mobile, Meteorology, other
	470-512 MHz 55.80% TV 14-20	470-512 MHz 49.00% Amateur, Radio Geolocation, Fixed, Mobile, Radiolocation	490-614 MHz 54.00% TV, DVB-T	470-862 MHz 66.00% Analogic and digital TV	470-766 MHz 40.02% Analogue TV, DVB-T
	512-608 MHz 55.70% TV 21-36	512-608 MHz 53.00% TV 21-36			
	608-698 MHz 55.50% TV 37-51	608-698 MHz 46.00% TV 37-51	614-790 MHz 9.00% TV		
	698-806 MHz 42.70% TV 52-69	698-806 MHz 44.00% TV 52-69	790-824 MHz 2.00% Mobile Data, PMR, Trunk, Fixed		766-880 MHz 12.30% Military, TV, DVB-T, Cordless, other
	806-902 MHz 54.80% Cell phone and SMR	806-902 MHz 44.00% Cell phone and SMR	824-890 MHz 28.00% Digital Cellular, Mobile Data, PMR, Trunk	862-870 MHz 34.00% Wireless microphones, RFID, SRDs	
	902-928 MHz 9.30% Unlicensed	902-928 MHz 10.00% Unlicensed	890-960 MHz 37.00% GSM900, RFID		880-960 MHz 46.80% GSM, E-GSM, Military
	928-960 MHz 29.60% Paging, SMS, Fixed, BX Aux, FMS	928-1000 MHz 19.00% Paging, SMS, Fixed, BX Aux, FMS		870-960 MHz 41.00% PMR, R-GSM 900, E-GSM 900	

Table 4.3: Spectrum Occupancy Table 960 – 3000 MHz

Chicago I (2005)		Chicago II (2007)		Singapore (2008)		Spain (2009)		Bucharest (2010)	
960-1240 MHz	3.60% IFF, TACAN, GPS, others	1000-1100 MHz	1.00%	960-1429 MHz	0.00% Aero Nav, Radar	960-1350 MHz	3.00% Aeronautical and satellite radio-location/radionav (military radars)	960-1525 MHz	2.36% Aero/Naval, Navigation, Radar, Military, Radio astronomy
1240-1300 MHz	0.00% Amateur	1100-1240 MHz	0.30%						
1300-1400 MHz	0.40% Aero Radar, military	1240-1300 MHz	0.30% Amateur			1350-1400 MHz	1.00% Ministry of Defense		
1400-1525 MHz	0.00% Space/Satellite, Fixed Mobile, Telemetry	1300-1400 MHz	0.00% Aero Radar, military	1429-1525 MHz	0.00% Fixed/Mobile, DAB	1400-1710 MHz	4.00% DCS 1800, DECT, Government		
1525-1710 MHz	0.00% Mobile Satellite, GPS, Meteorological	1400-1525 MHz	0.00% Space/Satellite, Fixed Mobile, Telemetry	1525-1710 MHz	0.00% Mobile Sat, Met Sat, Aero Nav			1525-1710 MHz	4.08% Satellite Mobile, Military, Meteorology
1710-1850 MHz	0.00% Fixed, Fixed Mobile	1525-1710 MHz	0.00% Mobile Satellite, GPS, Meteorological	1710-1880 MHz	20.00% GSM1800, Fixed	1710-1900 MHz	25.00% DCS 1800, DECT, Government	1710-1880 MHz	22.86% GSM 1800, other
1850-1990 MHz	42.90% PCS, Asyn, Iso	1710-1850 MHz	0.00% Fixed, Fixed Mobile	1880-2400 MHz	4.00% PCS, Asyn, Iso DECT, TDD, 3G, Fixed/Mobile, WBA	1900-2300 MHz	10.00% UMTS (TDD and FDD), MSS, fixed and mobile service	1880-2200 MHz	14.50% UMTS/IMT 2000, DECT, other
1990-2110 MHz	2.20% TV Aux	1850-1990 MHz	40.00% PCS, Asyn, Iso DECT, TDD, 3G, Fixed/Mobile, WBA	1990-2110 MHz	2.50% TV Aux				
2110-2200 MHz	0.20% Common Carriers, Private Companies, MD	1990-2110 MHz	2.50% Telemetry	2110-2200 MHz	0.50% Common Carriers, Private Companies, MD			2200-2400 MHz	2.89% SAP/SAB, Military
2200-2300 MHz	0.20% Space Operation, Fixed	2110-2200 MHz	0.50% Common Carriers, Private Companies, MD	2200-2300 MHz	0.00% Space Operation, Fixed				
2300-2360 MHz	19.90% Amateur, WCS, DARS	2300-2360 MHz	18.00% Amateur, WCS, DARS	2300-2360 MHz	18.00% Amateur, WCS, DARS				
2360-2390 MHz	0.00% Telemetry	2360-2390 MHz	2.50% Telemetry	2360-2390 MHz	2.50% Telemetry				
2390-2500 MHz	29.10% U-PCS, ISM (Unlicensed)	2390-2500 MHz	1.00% U-PCS, ISM (Unlicensed)	2400-2700 MHz	0.00% ISM, WBA			2400-2500 MHz	9.42% ISM, RFID, LAN, other
2500-2686 MHz	30.80% ITFS, MMDS	2500-2640 MHz	2.00% ITFS, MMDS	2500-2640 MHz	2.00% ITFS, MMDS			2500-2690 MHz	2.21% UMTS/IMT 2000, Military
2686-2900 MHz	2.20% Surveillance Radar	2640-2686 MHz	29.00% ITFS, MMDS	2686-2900 MHz	0.10% Surveillance Radar				
		2686-2900 MHz	0.10% Surveillance Radar	2700-3400 MHz	0.00% Aero Nav, Radiolocation	2690-3000 MHz	3.00% Satellite, radionavigation and radiolocation (military radars)	2690-3400 MHz	3.70% Military, Radar, Navigation, Meteorology, other

800 MHz GSM band, providing 2G cellular services, is located in this spectrum range and shows a lot of activity. An additional reason for the high activity in the lower frequency ranges is the bigger transmission range for these frequencies.

In the spectrum between 1000 and 3000 MHz, the measured activity is a lot less. Exceptions are the cellular (GSM and UMTS) bands. Note that most measurement campaigns even recorded very limited activity in the 2.4 GHz ISM band. This is mainly due to the locations of the measurement equipment which was mostly on rooftops where usually no WLAN networks or other devices operating in this band are active. This observation also reveals one of the problems of many of the wide-band measurement campaigns: they are not suitable to identify the spectrum usage in all frequency bands. In almost all of the measurement campaigns, the antenna is placed on the rooftop of high buildings. Whereas such a setup is suited to detect relatively long-range, outdoor spectrum usage (such as TV broadcasting, cellular systems), it is not suited to detect indoor, short-range spectrum usage (such as WLAN, Bluetooth, and other systems in the 2.4 GHz ISM band).

4.1.3 TV Band Measurements

The TV bands often received special attention in the above measurement campaigns. This is not least due to the FCC ruling opening the TV bands in the U.S. for secondary usage [47, 49]. Most of the investigations summarized above show detailed plots and results of the occupancy in the TV bands. The measurements indicate that in all locations some of the allocated TV bands are used while some are not. Which TV bands are occupied obviously depends on the location of the measurements.

In the “Chicago II” measurement campaign [11], e.g., the detailed occupancy of all TV bands is shown. It is interesting to observe that also the apparent temporal transmission failure on some TV bands was detected. Furthermore, the paper shows the different spectral signatures of analog (NTSC) and digital (ATSC) transmissions.

In the “Singapore” measurements [85] it is shown that some broadcasting stations stop broadcasting at night, opening possibilities for secondary usage. The measurements also suggest that in some TV bands only very weak signals are received. However, even in those bands the pilots are usually strong enough to be detected.

A measurement campaign analyzing the occupancy in the TV bands in Chengdu, China is presented in [67]. The usage of 14 TV bands of 8 MHz bandwidth each in the 694 to 806 MHz range is investigated. The results are pretty much in conformance with the previous results, namely that while some TV bands are detected to be constantly used or idle, there are some where broadcasting seems to be interrupted at night. In the results, also an outage of a couple of hours was observed in one TV band.

4.1.4 Cellular Measurements

Another set of spectrum bands which received a lot of attention in spectrum measurements are the cellular bands. The most popular standards used in these bands are the 2G standards GSM and CDMA-based IS-95, and the 3G standards UMTS and CDMA2000.

In some locations additional test licenses have already been given for the 4G LTE standard.

Although the cellular bands are said to be one of the most efficiently utilized bands, they have been very popular for spectrum occupancy measurements. One of the reasons for that is the global deployments and usage of these bands and the vast amount of knowledge and experience gathered with the technology, which lowers the initial costs for developing cognitive approaches in these bands. Another reason is that those technologies are pretty well understood, so spectrum measurements can be fairly easily analyzed to estimate the occupancy. Thus, most of the wide-band measurements presented above also put a special emphasis on the cellular bands.

In the “Chicago I” measurement [109], the 806 – 904 MHz is identified as the main cell phone band in Chicago, with a relatively high estimated duty cycle of 55%. An interesting observation of this measurement is that the measured occupancy is much higher in the downlink bands than in the uplink bands. This effect is also visible in most of the other investigations and can be explained as follows: (1) the power output of the base station is a lot higher than that of the cell phones, (2) the base station continuously emits signals on the control channels, resulting in a high measured occupancy, and (3) measurement antennas and base stations are located at rooftops while stations are usually used on the ground floor or indoors. [85, 101].

In addition to the measured difference in uplink / downlink occupancy, the “Spain” measurements [101] give details about the occupancy measured in the different 2G and 3G spectrum bands. From the measurements, the authors conclude that the uplink in the 1800 MHz band has an even lower occupancy than the uplink in the 900 MHz band, which is due to the lower transmission power used. This also applies for the UMTS bands (1900 and 2100 MHz bands), where the reason is the very low transmission power of the spread spectrum CDMA technique used for UMTS. The measurement results also suggest that at the time and location of the measurement campaign some of the UMTS bands were completely unused.

There are also several measurement campaigns specifically focusing on the cellular bands. During the soccer world-cup 2006 in Germany, e.g., a measurement campaign focusing on the cellular voice bands was carried out [69, 132]. The authors show the differences in measured spectrum occupancy in both the GSM and UMTS bands before, during, and after a match. Similarly, in [14] the spectral occupancy of a GSM network is analyzed and compared at a big event, at a location near the big event, and at a “regular” reference location. Based on the Cumulative Distribution Function (CDF) of the measured PSD, the authors classify the channels into “noise” (always low power), “traffic” (mixture of low and high power) and “beacon” channels (always high power).

4.2 Limitations of the Presented Spectrum Measurement Campaigns

From the measurement campaigns presented above it seems that there is an abundance of available spectrum throughout the whole radio spectrum. However, to which extent

does the measured occupancy correlate with the real occupancy? What additional information about the spectrum occupancy (such as variability) can be extracted from the measurements? In this section, we analyze the above measurement results trying to answer these questions.

4.2.1 Measured Occupancy vs. Real Occupancy

The most fundamental problem of all the measurement campaigns summarized above is that the measurement results only give an **estimation** of the spectrum occupancy. In none of the campaigns, the results are validated against the **real** occupancy.

Determining the Real Spectrum Occupancy

Knowing the real occupancy requires access to the PU network; however even with access to this data, determining the exact spectrum occupancy is challenging. In the case of TV broadcasting, e.g., having access to the PU network, it is pretty easy to determine whether or not a certain television tower broadcasts on a certain TV channel. However, whether or not a TV receiver can receive the broadcasted signal at a certain location depends not only on its distance to the television tower, but also on many other factors such as the terrain, exact location of the receiving antenna (indoors, outdoors), the sensitivity of the TV receiver, the weather, etc. So even if a TV receiver should theoretically be within the coverage range of the television tower, it might be that no signal can be received or that, e.g., reception is good on sunny days in contrast to rainy days.

To ensure that the primary communication is not disturbed, the real occupancy is, thus, usually estimated assuming the best case propagation of the primary signal (using an optimistic propagation model). The alternative would be to place a primary transmitter and receiver in every possible location of the operation range of the secondary network and experimentally determine whether or not communication is possible, which is not realizable in practice. The propagation model is used to determine the coverage area of the primary transmitter and whenever the primary transmitter is sending data, the respective spectrum is assumed to be occupied within the whole coverage area.

Assessing the Accuracy of the Occupancy Estimation

Having access to the PU network to determine the real occupancy is often very difficult, if not impossible. PUs are often not willing to grant access to such information. Assessing the accuracy of spectrum measurements by directly comparing them to the real occupancy is, thus, often not possible. But even without this tool of validation there are certain aspects of the measurement methodology and results which can be used to assess the accuracy of the occupancy estimation. In the following we point out some of those aspects found in the presented measurement campaigns.

In almost all campaigns investigated above, the occupancy estimation is only based on the results of one single spectrum sensor in one location. To be robust against shadowing

and fading, the occupancy estimation should be based on measurements concurrently done in multiple locations. If one of the sensors is in a deep fade, the other might still be able to correctly detect the PU. However the distance between the sensors should be limited, otherwise the real occupancy might already differ between the locations of the sensors.

Looking at the results of the New Zealand measurements [27], we can observe an example for the above mentioned problem. The campaign presents results for two measurement locations: one outdoors at the ninth floor and one indoors at the third floor of the same building. Although the measurements have not been done concurrently, we can assume that the results are comparable, since only long-term averages are shown. A general observation of the New Zealand measurements is that the signal strength of the outdoor measurements is 5 – 25 dB higher than in the indoor case, which is reasonable, since the indoor measurement location suffers from heavier shadowing due to the walls of the building. Comparing the estimated occupancy of the indoor and outdoor measurements shown in Table 4.1, shows that the indoor occupancy is always (except for aeronautical radio) lower than the outdoor occupancy. The question arises whether the occupancy is really lower at the indoor location, or whether the lower estimated occupancy is a measurement error due to the shadowed signal. In case of the GSM band, e.g., we would intuitively not expect the occupancy to have such significant differences as shown in Table 4.1 for locations only some 10 meters apart.

Another aspect, also discussed in some of the campaigns, is the difference between the measured occupancy of the uplink and downlink cellular bands. All above measurement campaigns have estimated a lower occupancy in the uplink bands compared to the downlink bands. As explained above, this is most likely due to the higher power emitted by the base stations and their locations compared to the measurement equipment. However, does this mean that the occupancy is really that much lower in the uplink bands, or is the measurement equipment just not able to reliably detect the uplink transmissions? Since the majority of the usage within the cellular bands is still telephony, we would usually expect a similar occupancy of the uplink and downlink bands. In some of the campaigns it is suggested to only look at the downlink occupancy and assume the same occupancy in the uplink band, which would be a conservative upper bound of the estimated occupancy. On the other hand, it could also be argued to *only* look at the uplink bands and use those for estimation of the occupancy. If no occupancy is detected in the uplink bands, all the receiving cell phones are most likely far away, such that they would not be disturbed by a secondary transmission.

4.2.2 Variability of Available Spectrum

Apart from the spectrum occupancy as estimated, e.g., by the average duty cycle (D_i) defined in Equation (4.2), also the *dynamics* of the primary usage are of great importance for Dynamic Spectrum Access (DSA) and Cognitive Radio (CR) research. D_i only provides information about the *average* occupancy, estimated at a single location over a certain time and frequency span. It can, thus, be used to assess whether there is spectrum available at all in the respective spectrum band within the observed time.

However, for secondary usage it is not only important *if* spectrum is available but also to which extent the availability *varies*.

The primary focus of most of the measurement campaigns summarized above is no doubt the identification of the parts of the spectrum which are largely underutilized and, thus, potentially attractive for secondary usage; consequently D_i is the main metric used. Since we are interested in the real-time estimations of PU activity needed to actually take advantage of the unused spectrum through secondary usage, we investigate the presented spectrum measurement campaigns with respect to the information on the variability of available spectrum. Note that we do not focus on the potential difference between the estimated and the real occupancy (as analyzed above) anymore. For the variability investigation we assume that the measurements provide at least to some extent accurate results. We investigate the variability of available spectrum in the three dimensions: frequency, space, and time.

Frequency Variability

The frequency variability is defined by the correlation of the primary usage in neighboring PU bands. Licenses for neighboring PU bands are often given to the same type of service and technology, having, thus, a potential for strong correlation of the primary usage. In the U.S. e.g., there are 56 6 MHz TV channels allocated in the 470 – 806 MHz spectrum band. The same applies to the cellular GSM bands, where there are, e.g., in Germany 174 GSM channels in the 880–915 MHz (uplink) and 925–960 MHz (downlink) spectrum band and 350 GSM channels in the 1710.2–1780.4 MHz (uplink) and 1805.2–1875.4 MHz (downlink) spectrum band. For UMTS the allocation in Germany is 1920 – 1980 MHz (uplink) and 2100 – 2170 MHz (downlink), resulting in 12 5 MHz channels.

Considering the limited bandwidth that can be processed by radio front-ends, occupancy correlation of neighboring PU bands can have a significant impact on the possibilities of secondary usage of the respective spectrum. For a CR system only able to operate in the TV spectrum band (470 – 806 MHz), e.g., the frequency variability is of eminent importance. If there is strong correlation between PU activity in the various bands, the CR system will either have an abundance of available spectrum or none at all. Note that the frequency variability is most interesting on a rather short time scale. Even if the average occupancy shows a strong correlation among neighboring PU bands, the usage may vary significantly on a short time scale, enabling to exploit the frequency variability for secondary usage.

The spectrum measurements summarized above only provide information on the frequency variability of available spectrum on a very large time scale. For the TV and cellular bands, e.g., almost all measurements show that some channels are used at a certain location, whereas others are not. Although some of the measurements record data with a relatively fine time granularity, details on a shorter time scale are not evaluated or published.

The correlation of the idle time durations and the duty cycle in the frequency domain are investigated in [24]. The authors use a time scale of 75 s. It is shown that correlation is very high in general, and specifically among similar services. E.g., the different

PU bands in the 900 MHz GSM band show a very high correlation. In the study of the IS-95 and GSM bands in New York by Kamakaris et al. [89], the correlation of the usage in neighboring bands is analyzed on a time scale of 7.5 min. Although a rather strong correlation is observed, it is shown that there is at least to some extent frequency variability which can be exploited by secondary usage.

Spatial Variability

For spatial variability of available spectrum we have to differentiate between the large scale (in the range of tens of kilometers) and the small scale (in the range of tens of meters) variation. On a large scale, the different wide-band measurements summarized in Tables 4.2 and 4.3 indicate that there is considerable variation of available spectrum among different locations. Although the different measurements have not been done at the same time, the long-term averages shown can be expected to be comparable. The tables also show that even the allocation of licensed users differs for the shown locations. A CR network, thus, has to be able to always adapt to the local spectrum occupancy and regulation. An essential principle of CRs is thus the policy driven operation as introduced in Section 2.1.3.

In [24] the spatial correlation of the spectral occupancy between four measurement locations (several tens of kilometers apart) is investigated. Whereas the occupancy averaged over several adjacent PU bands of the same service is shown to be significant, no spatial correlation of the occupancy of *individual* PU bands could be observed.

The more interesting case for secondary usage is the spatial variability of available spectrum on a small scale. Unfortunately, none of the measurements presented above gives information on the small scale spatial variability. To get information on the small scale spatial variability, multiple sensors are required to perform concurrent measurements in different locations (similar to the measurements presented in [24], but in a much smaller area).

Small scale spatial variability can have a big influence on secondary usage if the CR network is mobile. Having a high spatial variability means that the available spectrum could change significantly with the location of the CR network. The same actually applies if the PU network is mobile. Although the usage in time might be constant, the occupancy as experienced by the CR can vary significantly due to mobility. Note that the variability in space is *experienced* as variability in the time domain. A CR will observe a varying occupancy in time. The influence of mobility obviously depends on the speed: the faster the relative speed between primary transmitter and sensing CR, the higher the impact. On the other hand, the transmission power of the primary transmitter is also important: the smaller the transmission power, the higher the spatial variability.

Time Variability

The time variability of available spectrum has a big impact on the secondary usage of the spectrum. Obviously, a PU band which is available for a long consecutive period of time is much more attractive for secondary usage than a PU band which is only available for

short durations, assuming the same average duty cycle. A high time variability results in a high frequency of potential reconfigurations of the secondary communication and, thus, increases the challenges for secondary QoS support mechanisms.

As seen above, the *spatial* variability has an influence on the time variability of available spectrum. Obviously, the PU agility influences the time variability as well. A license owner frequently accessing its band (for short durations) results in a high time variability of the available spectrum. The PU agility has a big influence on the maximum interference time (t_{\max}), and thus on the design of spectrum sensing approaches: for highly agile PUs, t_{\max} is usually very short, whereas rather static PUs usually can tolerate a longer t_{\max} . Consider, e.g., TV broadcasting compared to wireless web traffic: a t_{\max} of several seconds might still be acceptable for TV broadcasting, whereas it could interfere with a whole http session in the case of web traffic. Thus, very agile PUs require a fast sensing process.

Most of the spectrum measurement campaigns analyzed in the last section have no or only very little information on the PU agility. Mostly only a duty cycle averaged over the whole observation period — as defined in Equation (4.2) — is shown, completely eliminating the time variability information. Some results provide at least some time variability information, such as the “Spain” measurements [101] for the 1800 MHz cellular bands: they show the occupancy difference between day and night, which have a regular 24-hour pattern.

For the TV bands, the PU agility can be observed from most of the measurement campaigns. However, this is simply due to the fact that they show a very static behavior. The band is either always occupied or always idle. The “Singapore” measurements [85] show some, very predictable, PU agility: some TV stations stop broadcasting at night.

The authors of [89] analyze the spectrum utilization in the New York cellular bands (IS-95 as well as GSM). The IS-95 signals are demodulated to determine the number of active Walsh codes (i.e., the number of ongoing calls). To determine the number of calls in the GSM bands, image processing of the spectrogram snapshots is used. The authors show details of the short term variation of the primary usage in the investigated bands by comparing the load of consecutive 7.5 min time slots. The investigations show that the variation is higher during the day (when also the general occupancy is higher) than during the night.

4.3 Ground Truth of Spectrum Availability

All the measurement campaigns presented so far provide only an estimation of the spectrum occupancy. Furthermore, only very limited information on the variability of available spectrum can be extracted from those campaigns. To get an idea of the *real* occupancy and to get precise and detailed information on the variability of the spectrum occupancy, we conducted a large-scale measurement-driven characterization of primary usage in cellular networks based on in-network measurements. With the conducted study we (i) show that there is unused spectrum even in these heavily used bands, and (ii) we provide detailed insights into the dynamics of the agile licensed users occupying the

cellular bands. The work presented in this section received the best paper award in the technical track of the IEEE DySPAN conference 2008 [170] and was also published in [171].

4.3.1 Measurement Methodology

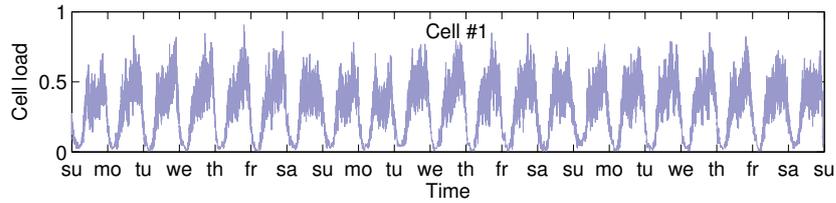
Our study is based on the analysis of a unique dataset consisting of call records collected *inside* a cellular network. This approach is completely different to the sensing-based measurements of the previous sections. The advantage is that very detailed usage information (down to the call level) is available and that the usage information is not subject to any sensing errors, i.e. represents the real usage by the PUs. Another advantage of our study is its scale — we are able to study usage at hundreds of base stations simultaneously. In contrast, sensing-based studies are usually based on only a few spectrum analyzers and, typically, have limited spatial resolution. Moreover, we are able to study the entire spectrum band used by a cellular operator. Sensing-based studies take time to “sweep” such a band and, hence, have to tradeoff the sampling frequency with the width of a band. The temporal diversity of our data is also large — we use measurements of tens of millions of calls over a period of three weeks. Finally, by looking at call records, we measure the “ground truth” as seen by the network, and, hence, are able to model call arrival processes as well as system capacity.

The dataset we use was collected from hundreds of cells² of a CDMA-based cellular operator. The data captured voice call information at those cells, which were all located in densely-populated urban areas of Northern California, over a period of three weeks. In particular, our dataset captured the start time, the duration, the initial and final cell of each call. Note that the call duration reflects the RF emission time of the data transmission for the call, i.e., the duration of time for which a data channel was assigned. This is precisely what is relevant for DSA questions. The start time of the call was measured with a resolution of several milliseconds. The duration was measured with a resolution of one millisecond. Overall, the collected data consists of tens of millions of calls and billions of minutes of talk time. To our knowledge, such a large-scale *network* viewpoint of spectrum usage has not been analyzed in any prior work.

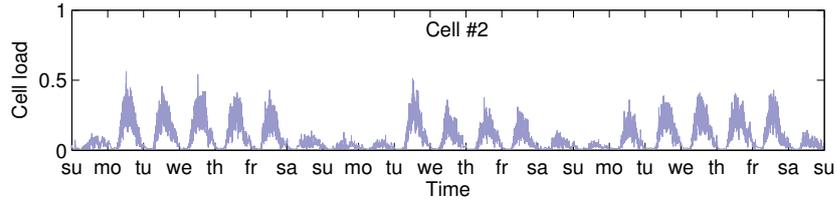
As with any measurement-based study, our dataset has certain limitations. We state these up-front since it is important to understand what our results capture and what they do not.

The first limitation of our dataset is its lack of full information on mobility. We were able to record only the initial and final cell of each call. Thus, we are unable to account for spectrum usage in the other cells that users may have visited during calls. To address the resulting incompleteness of information, we use two types of approximations. In the first approximation, we assign the entire call as having taken place in the initial cell. We use this approximation by default. In the second approximation, we assign the first (last) half of the call to the initial (final) cell. We refer to this as the *mobile* approximation. Throughout the paper, we provide results using both approximations and find that our

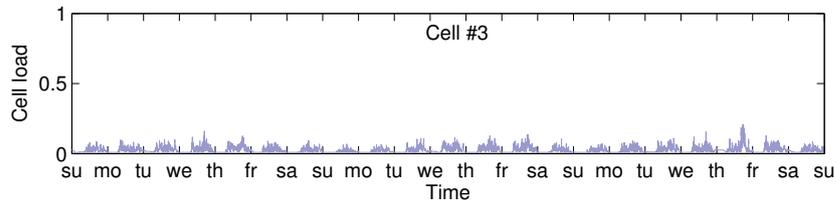
²We do not give the specific number for proprietary reasons.



(a) Cell 1: high load



(b) Cell 2: varying load



(c) Cell 3: low load

Figure 4.2: Normalized load of three different cells over 3 weeks. We plot the moving average of the load in each cell over one second.

conclusions do not change. These results indicate that the results are not sensitive to our approximations and would likely not change with full mobility information.

The second limitation relates to the cellular system from which we collected our dataset — a CDMA-based network. Without additional knowledge from the base stations, the precise CDMA system capacity cannot be easily calculated. Hence, we implicitly assume that each voice call uses the same portion of a cell capacity. This assumption, which is correct for TDMA-based systems like GSM, is obviously not precise for CDMA. Due to the critically important power control loop, individual CDMA calls may require different portions of the cell capacity, which cannot be easily expressed only in the number of calls. Nevertheless, since user calling behavior is unlikely to depend on the underlying technology, except under rare overload conditions, many aspects of our analysis are likely to apply to other cellular voice networks.

Using either of the aforementioned approximations, we compute the total number of ongoing calls in each cell during the entire time period of our study. To do so, we split the call records based on the cell. We create two records for each call — corresponding to the beginning and end time of each call. Then, we sort these records in order of their

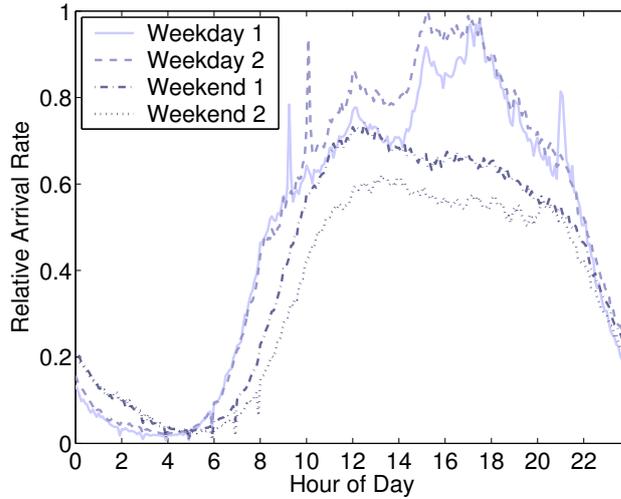


Figure 4.3: Distribution of system-wide average call arrival rates during four different days. The arrival rates are averaged over 5-minute slots.

time. We maintain a running count that is increased by $+1$ when a call begins and decreased by -1 when a call terminates.

4.3.2 Dynamics of Spectrum Availability

We plot the obtained “load” of three representative cells in Figure 4.2. For proprietary reasons, we normalize the values of load by a constant value such that only the relative change is seen. Cell 1 (Figure 4.2(a)) has low load only at night whereas cell 2 (Figure 4.2(b)) has low load during the weekends too (note that the second Monday in the observation period was a public holiday). Cell 3 (Figure 4.2(c)) always has low load (during both day and night). The plots show clearly that there are numerous periods of available spectrum also in cellular networks. For some cells the spectrum is even constantly available for reuse. However, the plots also show that spectrum usage varies widely over time *and* space. Since only a single cellular band was observed, no details on the frequency variability of the spectrum usage can be extracted from our measurements. In the following, we focus on the detailed investigation of the time variability of the spectrum usage. For details on the spatial variability, refer to [170].

The day/night dependence is exhibited system-wide as seen in Figure 4.3. Here, we ignore information about the individual cells to which calls are assigned and consider all calls as arriving to a single entity. For such a hypothetical system, we plot the normalized average call arrival rates during four different days. Figure 4.3 illustrates three key effects regarding the dynamics observed in the system. First, there are two distinct periods, which roughly correspond to day and night and have high and low arrival rates respectively. Moreover, the steepest change in arrival rates occurs in the morning and late in the evening, which correspond to the transition between the day

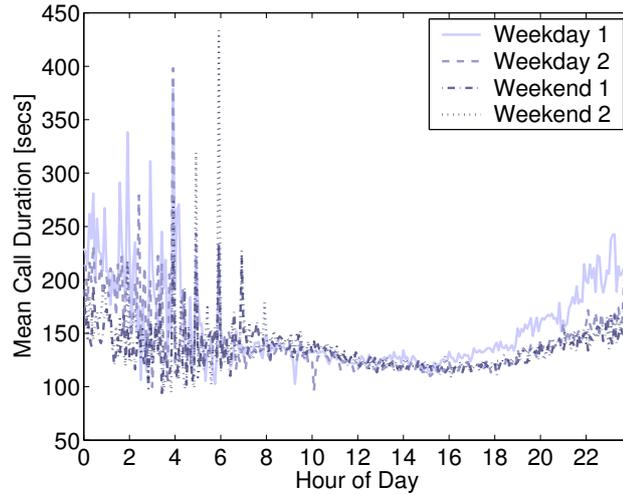


Figure 4.4: Distribution of average call duration over 5-minute periods during four different days. The large spikes during the mornings are due to small gaps in collection.

and night periods. Second, the system characteristics are unlikely to remain stationary at timescales beyond an hour. Except for the transition hours, the mean arrival rates do not vary significantly during an hour. Third, weekdays and weekends appear to show distinct trends.

We plot the average call durations as a function of time in Figure 4.4. From the figure, we find that the range of variability in mean call duration is much smaller than that of arrival rates. Note that there are a few large spikes in Figure 4.4. These are caused by a brief interruption in the data collection, which caused some short calls to not be recorded, thereby artificially inflating the mean duration of calls.

4.3.3 Primary User Agility

Secondary usage requires the availability of free spectrum. Assuming secondary users are immobile, the best scenario is the one in which free spectrum is always available for as long as possible in any given cell. In other words, variability in per-cell spectrum availability is not desirable. We quantify this variability by computing the variation in the load of each cell during each hour. We calculate the “average-case” variation using the standard deviation and the “worst-case” variation as the difference between maximum and minimum 1-minute load in a cell during each hour. We average these over all cells and plot them on an hour-of-day basis in Figure 4.5. As before, we normalize the metrics to a maximum variation for proprietary reasons. Notice that both metrics show the same trends. Not surprisingly, the variation is larger during the day, when the load is higher.

One of the primary requirements of DSA-based approaches is that secondary spectrum usage should not affect PUs. Hence, it is critical that CRs in cellular networks

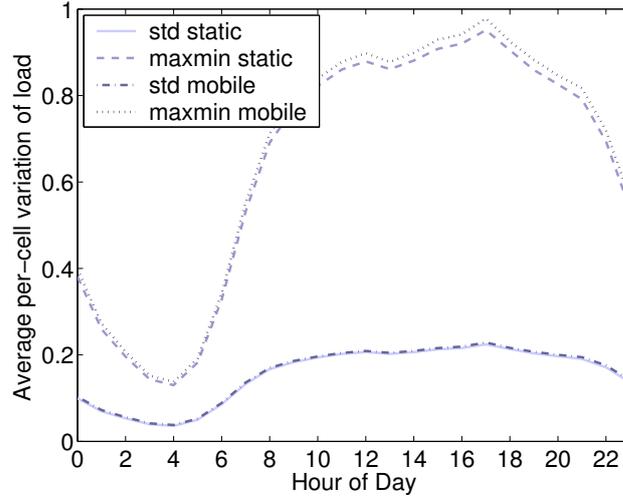


Figure 4.5: Average per-cell variation of load on an hour-of-day basis. We calculate the variation using the standard deviation and the difference between maximum and minimum.

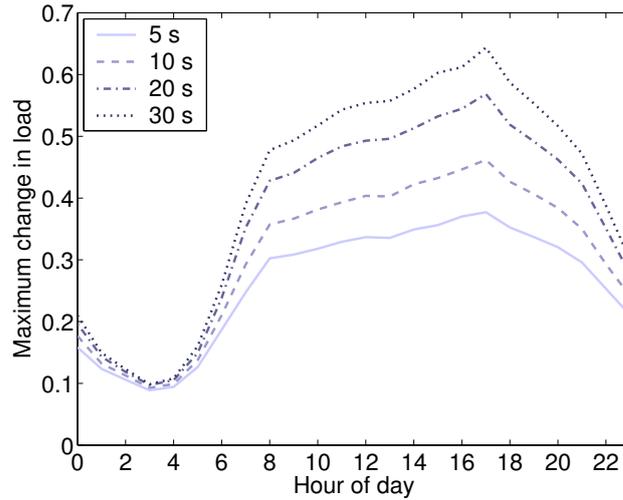


Figure 4.6: Maximum change in load, averaged across all cell sectors, plotted on an hourly basis. We use different time windows t_{\max} over which the maximum change is calculated.

frequently sense the spectrum and vacate it if a PU is detected. Also, since the available spectrum could change between two consecutive sensing periods, CRs must be aware of the extent of such short-term variations. As explained in the previous section, the PU agility has an influence on the maximum interference time (t_{\max}). Remember that t_{\max} specifies the time a PU band can be used for data transmission before it has to

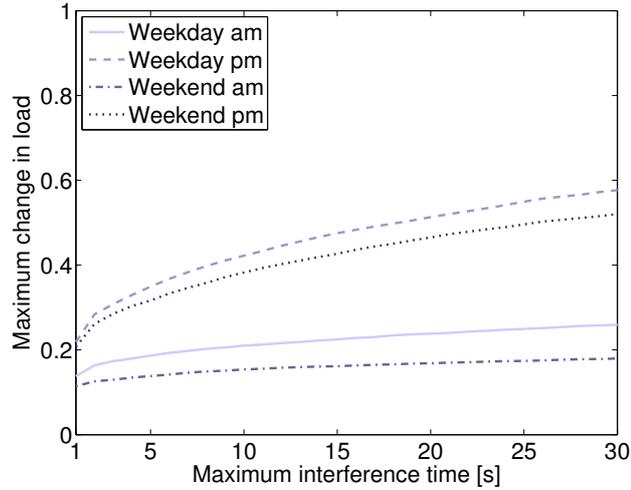


Figure 4.7: Maximum change in load, averaged across all sectors, plotted as a function of t_{\max} for 4 different hours.

be sensed for PU activity again. Using the data from our measurements, t_{\max} could even be dynamically adapted to the PU agility: for a high agility t_{\max} needs to be very short whereas a longer t_{\max} can be chosen for low agility of the PU. Figure 4.6 provides insights into this by plotting the maximum increase in load for different values of t_{\max} averaged over all cells. We plot the variation during a representative day of our dataset. The low variations at night are seen again. We see the peak variations in the late afternoon and a steep reduction thereafter. Notice also that the variation within a time span of $t_{\max} = 30$ s is often close to the variation within a time span of $t_{\max} = 5$ s (especially at night) and never more than twice. This indicates that 20 – 30 seconds may provide a better tradeoff between sensing overhead and the spectrum that CRs need to leave unoccupied for a sudden arrival of PUs.

We take a detailed look at the variation of t_{\max} for four representative hours in Figure 4.7. We see less variation during the weekend, possibly due to the reduced average load. We also see that during the AM hours, a small t_{\max} (1 – 2 seconds) does not pay off, since the maximum change in load increases only slightly. We found this to be true for all morning hours (before 10 AM). In the afternoon hours, however, there might be benefits from using a small t_{\max} .

4.3.4 Implications for Secondary Usage

Knowing the spectrum occupancy of a PU, or, more precisely, the dynamic change of the occupancy over time, is crucial for determining the degree to which secondary usage can be allowed, for example, as discussed in [84, 143]. First of all, the instantaneous occupancy sets an upper limit on the resources available for SUGs. Thus, our results in Figure 4.3 indicate that significant secondary usage is possible during the night until

almost 7 AM, regardless of the location. Additionally, in some locations, spectrum can become available during the day at weekends and even throughout the whole week. Knowing the future trends of occupancy further helps spectrum owners optimize their auction process without impairing PUs. For instance, if the primary spectrum occupancy tends to vary significantly (as can be observed in Figure 4.5 for the afternoon hours), secondary usage has to be allowed more conservatively, such that enough resources are available for new PUs. On the other hand, if the PU occupancy tends to decrease, spectrum can be rented more aggressively. Figure 4.5 highlights a significant challenge for cellular DSA: when there is less spectrum available, the availability is more variable, too. Hence, secondary spectrum usage has to be more conservative when less spectrum is available.

Being able to develop models for the PU agility based on the information evaluated above would be of great value for secondary usage of spectrum. They could be used by the PU to predict the amount of resources that can be leased to secondary users. On the other hand, models for PU agility are also required to develop and evaluate spectrum sensing and secondary QoS support concepts. In Appendix A we present a PU agility model developed based on the data presented above. The developed model is used to simulate primary usage in the investigation presented in Chapter 7 of this thesis.

4.4 Estimating the Variability of Available Spectrum

Knowing the ground truth of the real spectrum availability, the question arises whether and how this real spectrum availability can also be estimated through spectrum sensing. Inspired by the lack of validation for all the measurement campaigns summarized in Section 4.1 and the in-network measurement presented in Section 4.3, we conducted an in-network measurement *in parallel* to a spectrum sensing measurement [173]. The main results are summarized below.

4.4.1 Validating Spectrum Measurements Against the Ground Truth

To evaluate if sensed power yields information about the underlying primary usage, we conducted a large number of sensing measurements spread over time and multiple locations. In addition, we simultaneously recorded detailed information about primary usage inside the network. Thus, we were able to collect a large amount of unique data consisting of both measurements *and* ground truth. Our measurements were collected in the band used by a CDMA-based cellular operator. The detailed measurement methodology, results, and evaluation can be found in Appendix B. Here, we only briefly summarize the main conclusions.

Our measurement indicates that the estimation of the available spectrum in cellular bands is generally possible. However, there are a lot of challenges to overcome. One of the main challenges of the conducted experiment is that the PUs in our study use CDMA as a transmission technology. For CDMA, spectrum usage cannot be estimated using a binary metric (used or not used) but rather by a continuous metric (how many CDMA

codes are used). Another challenge we face is that we can only use a single spectrum sensor. As explained above, single sensor measurements are prone to estimation errors due to fading, etc.

Nonetheless, we gain a lot of insight into the spectrum sensing process. We show that it is possible to detect all calls in the primary network. The price, however, is that the estimation process has to be oversensitive, i.e., often declares a call to have started, although the ground truth data does not show such an event. In other words, to be able to keep the probability of false negatives (P_{fn}) low, a lot of false positives have to be accepted in the detection process.

These initial results motivate us to perform a more thorough analysis. Is this high probability of false positives (P_{fp}) a distinct phenomenon for the CDMA-based PU bands? Can the false positives be reduced using multiple sensors? We answer these questions in Chapter 5 of this thesis.

4.4.2 Influences on the Time Variability of Available Spectrum

False positives also have an impact on the time variability of available spectrum experienced by the CRs. A false positive will also be reported as a PU detection, i.e., the respective spectrum is reported to be occupied. A high probability of false positives, thus, results in an increase of the time variability of available spectrum.

To summarize, there are three effects which influence the time variability of available spectrum:

1. First of all the PU agility obviously impacts the time variability of available spectrum. The more agile a PU, the more often the availability changes.
2. As explained earlier, also the spatial variability in combination with mobility influences the time variability of available spectrum.
3. Last but not least, the quality of the sensing process, quantified by the probability of false positives, can have a big impact on the time variability of available spectrum.

These observations confirm our assumptions that CR systems need to be able to operate in environments with a rather high variability of available spectrum.

CHAPTER 5

Reliable Detection of Primary Users

This chapter addresses the reliable detection of Primary Users (PUs) through a sensing process. First, performance metrics are introduced followed by a feasibility study. The feasibility study, which is an extension of [166], analytically evaluates whether and under which circumstances reliable PU detection is possible for a simple, energy detection based sensing process. Next, diversity approaches for spectrum sensing are experimentally investigated. Specifically, the tradeoff between time and frequency diversity is exemplarily analyzed for selected scenarios. Finally, related work is presented and the chapter is concluded.

5.1 Performance Metrics

The performance metrics usually associated with the spectrum sensing process can be categorized into two classes. The first class, referred to as *reliability and quality* in this thesis, quantifies the result of the spectrum sensing process, i.e., the amount of spectrum *correctly* identified as available or occupied. The second class, referred to as *costs and complexity*, quantifies the means and required overhead to achieve the correct detection of the spectrum occupancy.

5.1.1 Reliability and Quality

The two main performance metrics commonly used for spectrum sensing are the probability of false negatives (P_{fn}) and the probability of false positives (P_{fp}). P_{fn} is a metric to quantify the *reliability* of the sensing process. It specifies the probability that a PU accessing its PU band is not detected by the sensing process. Complementary, P_{fp} specifies the probability that the sensing process reports the presence of a PU in a sensed PU band, although no PU is present in this band. P_{fp} is a metric to quantify the *quality* of the sensing process.

The accuracy and quality metrics defined above are subject to certain constraints. The detection process obviously has to make a decision on the state of an observed PU band (whether a PU is present or not) within a finite amount of time, which has a strict upper limit as explained below. Furthermore, also the sensing bandwidth (b_{sens}), i.e., the amount of spectrum used for the sensing process is obviously limited by the bandwidth of the PU band (b_{pu}), i.e., $b_{sens} \leq b_{pu}$.

The time available to make a decision on the presence of a PU is limited by the maximum interference time (t_{max}). The maximum interference time specifies the time limit, after which interference will be harmful for a licensed user. It is usually specified by regulation or the PU itself and can vary for different PUs. To assure the protection of the PU communication, the sensing process has to decide on the presence of a PU within t_{max} to enable the timely vacation of the PU band, if necessary.

The time required to detect a PU and to vacate the respective PU band is determined by the local sensing time (t_{sens}) and the distributed sensing time (t_{coop}).¹ The local sensing time specifies the time required by one sensor to locally decide whether a PU is present in the observed PU band or not. The distributed sensing time specifies the time required to exchange the local sensing results and come to a joint decision on the presence of the PU. The sum of both is strictly bounded by the maximum interference time ($t_{sens} + t_{coop} \leq t_{max}$), where $t_{coop} = 0$ if only a single sensor is considered and no distributed sensing is applied.

Note that b_{pu} and t_{max} are upper limits. A certain sensing process does not have to use the whole PU bandwidth for the detection process and can use a sensing time significantly shorter than the maximum interference time. In fact, the influence of vary-

¹The time required to physically stop transmission within a certain PU band (where a PU was detected) is negligible compared to t_{sens} and t_{coop} and, thus, not considered in this thesis.

ing t_{sens} and b_{sens} within the limits specified above can be quantified by the spectral overhead metric specified below.

5.1.2 Spectral Overhead

Whereas P_{fn} and P_{fp} are metrics to quantify the performance of the sensing process, additional metrics are needed to quantify its costs and complexity.

The *spectral overhead* is used as a metric to quantify the cost in the spectral domain. The spectral overhead quantifies the amount of spectrum which is used by the sensing process and, thus, blocked for other usage. It is mainly influenced by the different diversity approaches to spectrum sensing introduced in Section 2.3.1.

A fundamental challenge in spectrum sensing is that reliable detection is only possible if the spectrum being sensed is not used for data transmission by the CR system at the same time. Thus, the *local* spectral overhead of spectrum sensing is the product of the local sensing time (t_{sens}) and the sensing bandwidth (b_{sens}). Additionally, there is the *protocol* spectral overhead of distributed sensing. If multiple sensors collaborate in the sensing process, the sensing results have to be exchanged and a joint decision has to be made. To do so, protocols are required for exchanging the necessary information which requires spectral capacity and, thus, adds to the spectral overhead of the sensing process.

Furthermore, also the probability of false positives (P_{fp}) indirectly influences the spectral overhead. If idle spectrum is falsely declared as occupied by the sensing process, the respective spectrum cannot be used and, thus, results in less spectrum available for secondary data transmission.

5.2 Is Sensing-based Detection of Primary Users Possible?

Sensing-based detection of PUs is a challenging task, especially for PUs having a very agile usage pattern. Before going into more experimental investigations, this section revisits some of the information theoretic work on sensing-based PU detection to answer the question of whether and under which circumstances reliable PU detection based on spectrum sensing is possible.

5.2.1 Theoretical Limits of Reliable Primary User Detection

To assess the problem of sensing-based PU detection, we first consider the general information theoretical limits of reliable PU detection. Those investigations can be divided into results for a single sensor and results for multiple sensors cooperating in the sensing process (distributed sensing).

Single Sensor Detection

Reliable, sensing-based PU detection of weak PU signals is very challenging when using only one single sensor. In fact, given a target probability of false negatives (P_{fn}) and a

target probability of false positives (P_{fp}) to be met, Tandra and Sahai [146] have shown that detection of PU signals below a certain Signal-to-Noise Ratio (SNR) threshold (referred to as SNR_{wall}) is impossible using only a single, energy detection based sensor. This is due to the noise uncertainty, i.e., inaccuracies in receiver noise estimation.

Noise at the sensing receiver is the aggregation of many different processes, including device-level noise (non-linearity of chip components, thermal noise) and interference from other CRs transmitting during the sensing process. Whereas device-level noise uncertainty cannot be reduced, uncertainty due to other CR transmissions can be avoided by coordinating the sensing times among neighboring CRs [136]. The idea is to establish a “no-talk” radius within which no transmission is allowed during the sensing process. While this approach can significantly reduce the noise uncertainty, it cannot eliminate it completely. Note that so far an ideal channel with an infinite coherence time and coherence bandwidth is assumed, i.e., the received signal does not suffer any fading losses. Introducing fading would require the sensing receiver to be able to detect signals at even lower SNRs to account for potential fading losses.

In [136] it is also shown that — theoretically — noise and interference uncertainty can be eliminated using coherent detection. Thus, reliable PU detection is theoretically possible even without the need of coordinating the sensing times of neighboring CRs. However, Sahai et al. [136] also show that, in practice, the gains from coherent signal processing are limited by the coherence time (and coherence bandwidth) of the channel, since coherent detection can only be performed for coherent signals. So, in order to overcome the SNR_{wall} the coherence bandwidth (b_{coh}) has to be bigger than the PU bandwidth and an infinite coherence time (t_{coh}) of the channel would be needed. But even assuming such an ideal channel, the achievable gains from coherent processing are limited by the local sensing time (t_{sens}) which has to be strictly smaller than the maximum interference time (t_{max}). So, in practice, coherent detection can lower the SNR_{wall} and reduce the requirements on sensing coordination, but cannot completely eliminate it.

Summarizing, by using only one single sensor, reliable sensing-based detection of PU signals below a certain SNR_{wall} is impossible, even if a perfect channel is assumed. While this result is independent of the complexity of the sensing algorithms, the SNR_{wall} can be improved up to a certain level by using coherent detection. Furthermore, coordination of the sensing times (establishment of a no-talk region) can improve the detection. However, despite the potential improvement of the SNR_{wall} through coherent detection and coordination of the sensing times, the fundamental limit remains. When using only one single sensor, reliable detection of PUs below the SNR_{wall} is impossible.

Cooperative Detection

Seeing the discussion for a single sensor above, the question arises of why a CR needs to be able to detect very weak signals close to, or below the SNR_{wall} . The reason is that the CR needs to account for shadowing and fading losses of the primary signal. In order to be able to also detect the PU under worst case conditions, the CR has to be able to detect signals at very low SNR values to take fading and shadowing losses into account.

Cooperative spectrum sensing performed by multiple CRs can overcome the need to detect signals at very low SNR values by exploiting the spatial diversity of the measurements [111, 120, 156]. In cooperative spectrum sensing, each CR locally performs spectrum sensing and the sensing results are merged to achieve a joint verdict. As explained in Section 2.3.1, merging of the sensing results and the joint verdict can be based on raw measurement data (soft decision combining) or on 1-bit decisions of the individual CRs (hard decision combining).

Let us consider as an example the frequently followed approach of logically or'ing 1-bit decisions of the individual CRs [111, 156]. Assuming m CRs cooperating in the sensing process, the joint probability of false negatives in this case can be computed as

$$\overline{P_{\text{fn}}(m)} = (P_{\text{fn}})^m , \quad (5.1)$$

thus reducing the P_{fn} requirements of the individual CRs. The joint probability of false positives in this case can be computed as

$$\overline{P_{\text{fp}}(m)} = 1 - (1 - P_{\text{fp}})^m , \quad (5.2)$$

resulting into an increased probability of false positives of the SUG, compared to the probability of false positives of the individual sensors. For the discussion of other 1-bit merging strategies refer to [120]. The results of all these investigations show that several tens of cooperating CRs are required in order to achieve acceptable false negative and false positive rates [111, 120, 156].

Summarizing, the fundamental limit of one sensor can be overcome by combining the sensing results from multiple sensors. Having a sufficient number of sensing CRs, the signal levels a single CR needs to be able to detect can be pushed to the nominal path-loss of the primary signal, i.e., a spectrum sensor does not need to be able to detect very weak signals due to deep fades anymore. Thus, reliable detection of PUs based on spectrum sensing is possible using spatial diversity. However, to achieve both a low probability of false negatives of the SUG ($\overline{P_{\text{fn}}}$) and a low probability of false positives of the SUG ($\overline{P_{\text{fp}}}$), a lot of CRs have to cooperate in the sensing process.

5.2.2 Limits of Energy-based Detection

Energy-based detection, mainly due to its simplicity, is a very popular approach for spectrum sensing. First of all, no details about the primary signal, such as transmission technique, modulation used, or other characteristics such as pilot tones, needs to be known to the sensing CR. The detection is solely based on the energy received from the primary signal. Thus, the energy detector can be used for the detection of *any* PU. The only knowledge required is the frequency range the PU is operating in and the detection sensitivity requirements (i.e., which level of interference is considered to be harmful). Another advantage of energy-based detection are the low hardware requirements. Generally, any receiver can be used as an energy detector. The quality of the detection result obviously depends on the sensitivity of the hardware, but also very low cost, low complexity receivers can be used.

Due to the above mentioned properties, the energy detector is an ideal candidate for the CR system design to be developed in this thesis: it is cheap and simple and has low complexity, but can also be used with more sensitive, high precision hardware. Based on the observation that sensing-based PU detection is possible in general, we investigate below whether this is also true for the specific case of energy detection.

The Energy Detector

We follow the principal functionality of the energy detector as defined by Urkowitz [150], which we summarize below for convenience. The goal of the energy detector is to distinguish between the following hypotheses:

$$\begin{aligned}\mathcal{H}_0 : y(t) &= n(t) \\ \mathcal{H}_1 : y(t) &= s(t) + n(t) ,\end{aligned}\tag{5.3}$$

where $y(t)$ is the signal received by the energy detector, $n(t)$ is the noise and $s(t)$ is the PU signal to be detected. \mathcal{H}_0 means the PU is absent and \mathcal{H}_1 that the PU is present. A possible test statistic for the energy detector using the energy received within a time interval of t_{sens} seconds is

$$T = \frac{1}{t_{\text{sens}}} \int_0^{t_{\text{sens}}} y^2(t) dt ,\tag{5.4}$$

and the energy detector would decide on the presence of a PU based on a detection threshold τ as shown in Equation (5.5).

$$\begin{aligned}T > \tau & \text{ PU present } (\mathcal{H}_1) \\ T \leq \tau & \text{ PU absent } (\mathcal{H}_0)\end{aligned}\tag{5.5}$$

Following the line of argument in [150], the test statistic of Equation (5.4) is well approximated by sampling the signal at Nyquist such that $2 \cdot b_{\text{pu}} \cdot t_{\text{sens}}$ micro-samples are required. More generally, assuming that b_{sens} is the sensing bandwidth, the number of micro-samples used for energy detection computes to $N_s = 2 \cdot b_{\text{sens}} \cdot t_{\text{sens}}$. The resulting test statistic is shown in Equation (5.6).

$$T = \frac{1}{N_s} \sum_{i=1}^{N_s} y^2[i] \quad \begin{array}{ll} y[i] = n[i] & : \mathcal{H}_0 \\ y[i] = s[i] + n[i] & : \mathcal{H}_1 \end{array}\tag{5.6}$$

The Probability Density Function (PDF) of this test statistic follows a chi-square distribution with N_s degrees of freedom in case no signal is present ($y[i] = n[i]$) and a non-central chi-square distribution with N_s degrees of freedom and a non-centrality parameter γ being the received SNR in case the PU signal is present ($y[i] = s[i] + n[i]$) [150]. For a large number of micro-samples used for energy detection (N_s) the PDF can be approximated as a Gaussian distribution [150] using the central limit theorem [128]. In this thesis, we follow the definitions of Liang et al. [99], who assume circularly symmetric complex Gaussian noise with zero mean and variance σ_n^2 , and a Phase-Shift Keying (PSK)

modulated PU signal whose samples $s[n]$ are i.i.d. with zero mean and variance σ_s^2 . For this case, P_{fn} and P_{fp} can be calculated as shown in Equations (5.7) and (5.8), where $\gamma = \sigma_s^2/\sigma_n^2$ is the received SNR and $Q(x)$ is the Q-function. For details on the derivation of the formulas refer to [98, 99].

$$P_{\text{fp}} = Q\left(\left(\frac{\tau}{\sigma_n^2} - 1\right) \sqrt{N_s}\right) \quad (5.7)$$

$$P_{\text{fn}} = 1 - Q\left(\left(\frac{\tau}{\sigma_n^2} - \gamma - 1\right) \sqrt{\frac{N_s}{2\gamma + 1}}\right) \quad (5.8)$$

Single Sensor Detection

In the simplest case, we can use the energy detector defined above to perform energy detection using a single CR for spectrum sensing. As long as an ideal channel is assumed and the signal does not suffer any losses due to fading, this is the optimal energy-based detection possible.

However, let us consider the more general case of a non-ideal wireless channel. We have explained in Section 2.3.1 that time and frequency diversity can be used to increase the accuracy of the sensing process and combat the effects of fading. This diversity can also be exploited for the simple case of energy detection. To do so, we follow the approach of Liang et al. [99], who define a soft-decision combining approach for the exploitation of time diversity, which we extend to also include frequency diversity. The local sensing time (t_{sens}) and sensing bandwidth (b_{sens}) are split up into mini-blocks, which are defined by the coherence time (t_{coh}) and the coherence bandwidth (b_{coh}) of the wireless channel. The number of mini-blocks, i.e., the number of sensing samples (M) used for the soft-decision diversity combining can be calculated as

$$M = \left\lfloor \frac{b_{\text{sens}}}{b_{\text{coh}}} \right\rfloor \cdot \left\lfloor \frac{t_{\text{sens}}}{t_{\text{coh}}} \right\rfloor. \quad (5.9)$$

Following the approach taken in [99], we define the hypothesis for the k^{th} mini-block as

$$\begin{aligned} \mathcal{H}_0 : y_k[i] &= n_k[i] \\ \mathcal{H}_1 : y_k[i] &= h_k \cdot s_k[i] + n_k[i] \end{aligned} \quad k \in M, \quad (5.10)$$

where h_k are zero mean, unit variance complex Gaussian channel coefficients characterizing the fading. Consequently, the received signal is constant within each mini-block and received signals of different mini-blocks are i.i.d. We now perform energy detection in each of the M mini-blocks and combine the results using data fusion (soft decision combining) as proposed in [99]. The test statistic for the final decision in this case can be calculated as

$$T = \sum_{k=1}^M \left(\frac{g_k}{N_s} \sum_{i=1}^{N_s} |y_k[i]|^2 \right), \quad (5.11)$$

where the term in parentheses is the test statistic of the k^{th} mini-block and g_k is a weighting factor. Note that the number of micro-samples for each of these mini-slots reduces to $N_s = 2 \cdot b_{\text{coh}} \cdot t_{\text{coh}}$. For $g_k = 1/\sqrt{M}$ (unknown channel coefficients) P_{fn} and P_{fp} can be calculated as

$$P_{\text{fn}} = 1 - Q \left(\frac{1}{\beta} \left(Q^{-1}(P_{\text{fp}}) - \gamma \sqrt{N_s} \sum_{k=1}^M |h_k|^2 \right) \right) \quad (5.12)$$

and

$$P_{\text{fp}} = Q \left(\beta Q^{-1}(1 - P_{\text{fn}}) + \gamma \sqrt{N_s} \sum_{k=1}^M |h_k|^2 \right), \quad (5.13)$$

where

$$\beta = \sqrt{1 + \frac{2\gamma}{M} \sum_{k=1}^M |h_k|^2}. \quad (5.14)$$

For details on the derivation of P_{fn} and P_{fp} refer to [99].

Let us consider as an example the energy-based detection of PU signals in the TV bands in the U.S., i.e., a TV channel bandwidth of $b_{\text{pu}} = 6$ MHz with a carrier frequency of $f_c = 600$ MHz and a maximum interference time of $t_{\text{max}} = 2$ s. Assuming a suburban scenario and an average velocity of $v = 1$ m/s, the coherence bandwidth can be estimated as $b_{\text{coh}} = 600$ kHz and the coherence time as $t_{\text{coh}} = 100$ ms (for details on the calculations refer to Section 2.3.1). We choose $b_{\text{sens}} = b_{\text{pu}}$, i.e., the whole PU band is used for the sensing process, making maximum use of the available frequency diversity. Let us further assume detection in the very low SNR regime ($\gamma = -37$ dB). The reasons for choosing such a low SNR are on the one hand that CRs also need to detect very weak PU signals to account for fading effects as explained above and, on the other hand, that we are interested in the performance of low cost, low accuracy sensors, which typically will have a high device noise, further reducing the SNR. For the assumed scenario we generate the Receiver Operating Characteristic (ROC) shown in Figure 5.1 using Monte-Carlo simulation and Equations (5.12) and (5.13). For the calculation of the “no diversity” case we use Equations (5.12) and (5.13) with $N_s = 2 \cdot b_{\text{pu}} \cdot t_{\text{sens}}$ and $M = 1$.

There are two general observation we can make in Figure 5.1, which plots P_{fn} over P_{fp} : the performance of the energy detector increases with an increased local sensing time and the diversity approach using soft decision combining performs much better than the approach without diversity. The first observation is to be expected: sampling the PU signal for a longer time results in more collected energy from the signal, which increases the detection performance. The superiority of the diversity detector is not that intuitive, but can be — conceptually — explained as follows. The total number of micro-samples (and, thus, the collected energy) is the same for the diversity and the non-diversity detector. For the diversity detector, the total energy is divided into M smaller sensing samples (collected in the M mini-blocks), whereas the non-diversity detector only uses one big sensing sample. The detection performance of one small sensing sample of the

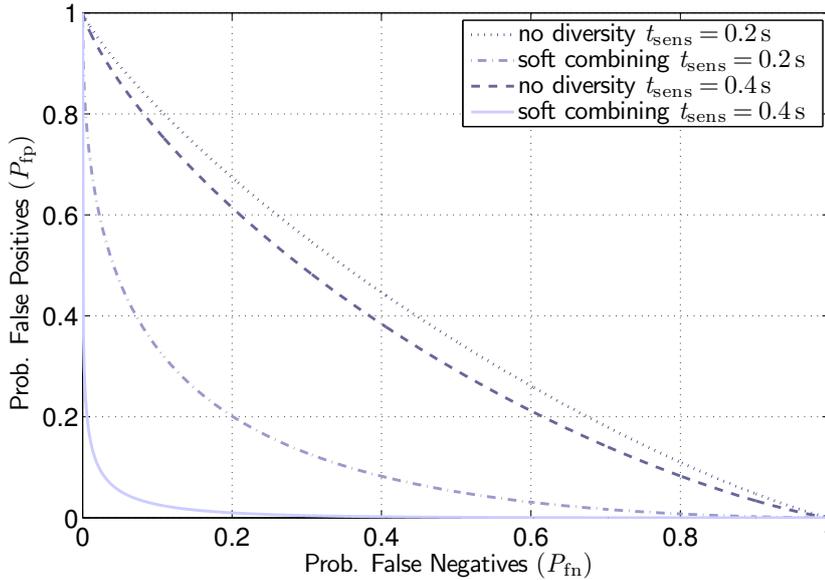


Figure 5.1: Receiver Operating Characteristic (ROC) for a single sensor

diversity detector is worse than the detection of the one big sensing sample of the non-diversity detector (this is in conformance with the first observation: more energy results in better performance). However, the small sensing samples of the diversity detector are combined again into a joint decision, exploiting the diversity of the sensing samples. From the example in Figure 5.1 we can conclude that the gain achieved through diversity is bigger than the loss due to the smaller amount of energy within the mini-blocks.

While the exploitation of diversity greatly improves the performance of the energy detector, there are certain limitations that have to be kept in mind. First of all the total time and bandwidth that can be used for the detection process is strictly limited by the PU bandwidth (b_{pu}) and the maximum interference time (t_{max}), which also limits the amount of signal energy that can be used for the detection. In practice these limits will be even stricter. Remember that no data transmission can be performed in the spectrum being sensed. So, in order to be able to also perform data transmission, the local sensing time and sensing bandwidth should be restricted to values well below the theoretical maximum. Second, the gains achievable through diversity are also limited. Obviously t_{coh} and b_{coh} put a limit on the achievable diversity. But even for very small t_{coh} and b_{coh} , there is a limit on the achievable gain through diversity: the mini-block used for the individual sensing samples will get smaller and smaller and with this also the energy for one sensing sample will become very small. At some point, the loss due to the smaller amount of energy used for the sensing sample will be bigger than the gain achieved through diversity.

The above discussion and the exemplary results in Figure 5.1 confirm the fundamental limits established in Section 5.2.1: achieving a low P_{fp} and P_{fn} using only one single

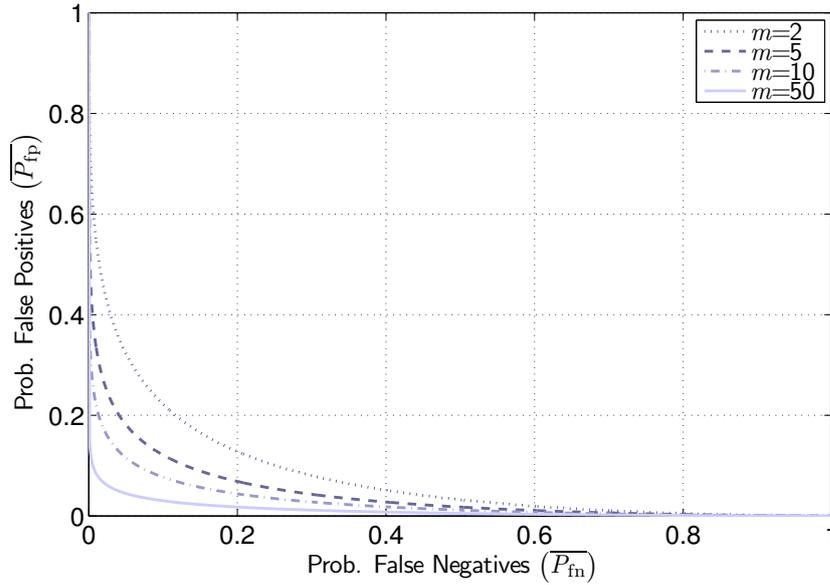


Figure 5.2: Receiver Operating Characteristic (ROC) for the whole SUG

sensor is impossible in a very low SNR regime.

Cooperative Detection

As indicated above, distributed sensing can help to overcome the fundamental limits of spectrum sensing with only one single sensor. Following the usually applied approach of logically or'ing the individual sensing results of a SUG where CRs cooperate in the sensing process, the probability of false negatives of the SUG (\overline{P}_{fn}) and the probability of false positives of the SUG (\overline{P}_{fp}) can be computed as a function of the number of cooperating sensors (m) (see Equations (5.1) and (5.2)). Following the example defined for the single sensor case and assuming a fixed local sensing time of the individual sensors of $t_{sens} = 0.2$ s results in the ROC shown in Figure 5.2.

The figure shows that spatial diversity increases the performance of the detection process: the more sensors cooperate, the lower \overline{P}_{fn} and \overline{P}_{fp} values can be achieved. The figure also clearly shows, in conformance with earlier results [111, 120] that in order to have both a low \overline{P}_{fn} and \overline{P}_{fp} a lot of sensors are needed. In fact, for the example shown in Figure 5.2, even using $m = 50$ sensors, the best we can achieve is a probability of about 5% for both \overline{P}_{fn} and \overline{P}_{fp} .

Energy Detection Limits

Looking at the single sensor results shows that relatively long local sensing times are required for the reliable detection of weak PU signals. However, such long sensing times are often not possible due to the restrictions imposed by the maximum interference

time (t_{\max}). Especially in case of very agile PUs, t_{\max} will usually be very small, resulting in a very limited performance gain that can be achieved through time diversity. Using cooperative sensing strategies, the local sensing time of the individual sensors can be kept pretty short. However, in order to have a reliable enough sensing process, a lot of sensors are required instead.

These results are a barrier for the deployment of initial CR systems. To enable early rollout and use of CR systems, very small groups of CRs have to be able to operate independently. In fact, the minimum number of CRs to form a SUG is two, i.e., one pair of CRs wanting to communicate. If tens of sensors are required to ensure reliable sensing, the only possibility for such small SUGs to operate would be to have additionally deployed sensing infrastructure which can help in the sensing process. Such a sensing infrastructure would have to be deployed, which requires big investments and, thus, is a barrier for fast and simple deployment of initial CR systems.

From these results it seems that simple energy-based sensing cannot be used for reliable PU detection. However, even using more complex sensing approaches based, e.g., on coherent detection does not overcome the fundamental challenge. Only having a very small number of cooperating CRs makes reliable spectrum sensing very hard. Furthermore, coherent detection requires expensive sensing hardware and complex algorithms, which creates a barrier for early CR deployments.

5.2.3 Reliable, Sensing-based Detection of Primary Users is Possible!

Despite the limitations and requirements elaborated above, we still believe that reliable PU detection through spectrum sensing based on energy detection is possible. However, to do so requires a fundamental change in the evaluation of the spectrum sensing metrics.

The reliable detection of PUs is the most important enabling metric for secondary spectrum usage. As outlined in the thesis approach, it is a hard constraint, which has to be met under all circumstances. In terms of spectrum sensing metrics, reliable PU detection is quantified by the probability of false negatives of the SUG ($\overline{P_{\text{fn}}}$). In order to assure reliable PU protection, $\overline{P_{\text{fn}}}$ has to be below a very strict, small threshold.

The probability of false positives of the SUG ($\overline{P_{\text{fp}}}$), in contrast, has no influence on the reliable PU protection. Thus, together with the spectral overhead of the sensing process, P_{fp} is a soft constraint which quantifies the *quality* of the sensing process but does not influence the *reliability* of the PU detection.

Based on this fundamental observation, the results in Figure 5.2 show that it is indeed possible to achieve a low $\overline{P_{\text{fn}}}$ with only very few sensors. However, for the results shown in the figure, a target probability of false negatives of the SUG of $\overline{P_{\text{fn}}} = 0.01$ would result in a probability of false positives of the SUG of $\overline{P_{\text{fp}}} = 0.33$ in case of $m = 5$ sensors and $\overline{P_{\text{fp}}} = 0.55$ in case of $m = 2$ sensors. For $\overline{P_{\text{fn}}} = 0.001$ the probability of false positives of the SUG would even rise to $\overline{P_{\text{fp}}} = 0.54$ in case of $m = 5$ sensors and $\overline{P_{\text{fp}}} = 0.75$ in case of $m = 2$ sensors. Thus, potentially an excessive number of false positives has to be accepted to assure the reliable detection of the PU. Those false positives require very frequent reconfigurations of the secondary communication, which is a major challenge for SUGs operating under these extreme circumstances. A potential CR system design

being able to tolerate such excessive false positives and still provide some secondary QoS is introduced and evaluated in Chapter 6 and Chapter 7, respectively.

5.3 Achievable Diversity in Spectrum Sensing

The above discussion shows that diversity in spectrum sensing has a significant impact on the spectrum sensing results. In this section, we quantify the diversity achievable for three different diversity approaches: time, frequency, and spatial diversity. Diversity in the local spectrum sensing process can be quantified by the number of sensing samples (M); diversity in the distributed sensing process by the number of cooperating sensors (m). In both cases, the samples have to be i.i.d. in order to add to the diversity of the measurement result. Following the model developed in the previous section, i.e., ensuring a low $\overline{P_{fn}}$ to assure reliable PU protection, there is a tradeoff between the amount of diversity used and the probability of false positives of the SUG ($\overline{P_{fp}}$): increasing the diversity decreases $\overline{P_{fp}}$ and vice versa.

Diversity in time and in frequency have a direct influence on the spectral overhead of the sensing process: they define the amount of time a certain bandwidth is blocked for the sensing process. Increasing the quality of the local sensing process (decreasing P_{fp}) through time and frequency diversity, thus, automatically results in a larger spectral overhead. Using spatial diversity requires spectral overhead not only for the local sensing process of the individual CRs but also for the exchange of the sensing data. In addition, the decision making process takes time and introduces an additional delay. Obviously, the spectral overhead for distributed sensing depends on m . The above tradeoff between diversity and $\overline{P_{fp}}$ can, thus, be translated into a tradeoff between the spectral overhead (required by diversity) and $\overline{P_{fp}}$. Remember that $\overline{P_{fp}}$ also influences the spectral overhead: idle spectrum falsely declared as occupied cannot be used for secondary communication. This spectral overhead tradeoff is analyzed in Chapter 8 of this thesis.

5.3.1 Experimental Measurements in the 1.9 GHz LTE bands

To verify the theoretical models for diversity described in Section 2.3.1, we have analyzed traces of a measurement campaign conducted in the 1.9 GHz LTE bands in France [87, 88]. The measurements were conducted using the Eurecom MIMO Openair Sounder (EMOS) [38].

The Measurement Traces

The presented diversity investigation is based on three measurements (labeled “meas 1, 2, & 4”), all conducted in the sub-urban Sophia Antipolis area, near Cannes in southern France. In all measurements, a base station, located at the rooftop of the Eurecom building, continuously sends an LTE OFDM modulated sounding sequence using 4 transmit antennas (2 transmit antennas in measurement 4). The terminals are laptops with Eurecom’s dual-RF CardBus/PCMCIA OpenAirInterface cards, which have two receive antennas. The recorded traces contain the estimated MIMO channel

values, i.e., a sub-channel time matrix for each transmit / receive antenna pair. The time granularity is 2.667 ms, which is the length of one transmit frame. The total bandwidth is 4.8 MHz and the frequency resolution depends on the number of transmit and receive antennas and is 120 kHz (40 sub-channels) in measurement 1 and 2, and 60 kHz (80 sub-channels) in measurement 4.

In measurement 1, 4 terminals are located in 4 different cars, which are driving around the area with no fixed routes and a velocity between 0 and 50 km/h. There is changing line-of-sight (LOS) and non-line-of-sight (NLOS) propagation between the transmitter and the receivers. In measurement 2, the 4 terminals are also located in 4 different cars, but this time the cars are parked close together in a parking lot in NLOS to the transmitter. Details on the above two measurements can be found in [88]. In measurement 4, the transmitter only uses 2 transmit antennas. There are two persons walking close together with one terminal each in LOS to the transmitter. These measurements is described in detail in [87].

Estimating Diversity

We first convert the channel sounder traces into fading profiles. Since we are only interested in the correlation of the *changes* in received power, we zero the mean by subtracting the average over all sub-channels in the frequency domain and the moving average of the next 100 frames in the time domain.

The Person product-moment correlation coefficient ($\rho_{x,y}$) is a commonly used metric to decide on coherence (or diversity) of two variables. It is defined as the covariance of the two variables divided by the product of their standard deviations as shown in Equation (5.15).

$$\rho_{x,y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \quad (5.15)$$

There are various definitions in the literature for the threshold of the correlation coefficient that defines coherence. In our work we follow the definition proposed in [63], in which two variables exceeding a correlation coefficient of 0.9 are said to be coherent.

We calculate the correlation coefficient for each transmit / receive antenna-pair separately and afterwards average over all antenna pairs to estimate the time and frequency diversity. We do this for every terminal separately. To estimate the spatial diversity, the measurements of the individual terminals are time-aligned and antenna pair and sub-channel values are averaged to get one fading profile per user in the time domain. We then calculate the correlation coefficients of these per-user fading profiles.

Evaluation Results

We start by looking at the frequency diversity of measurement 1. Figure 5.3 shows the correlation coefficients over frequency for all 4 users in measurement 1. In the top left corner of Figure 5.3, a zoom into the interesting region for a correlation coefficient of 0.9 is shown.

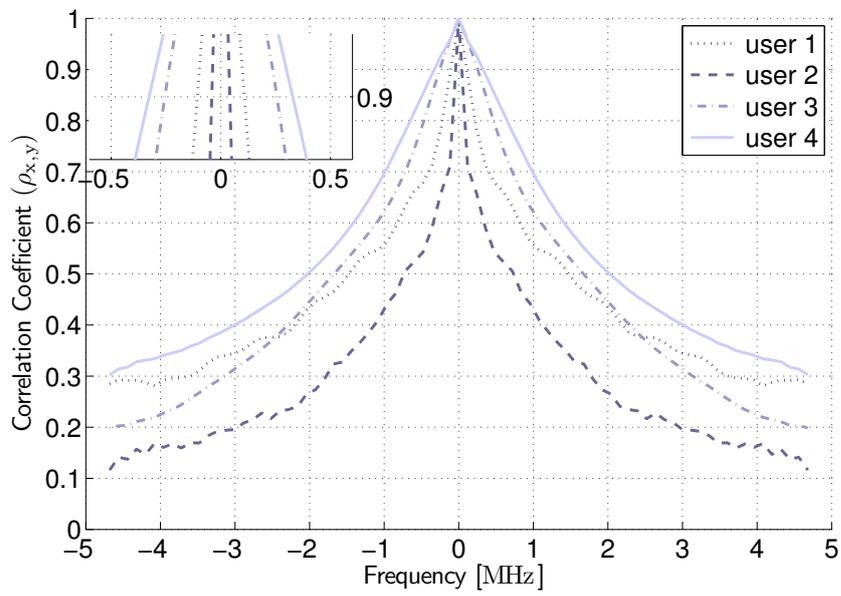


Figure 5.3: Frequency diversity in measurement 1

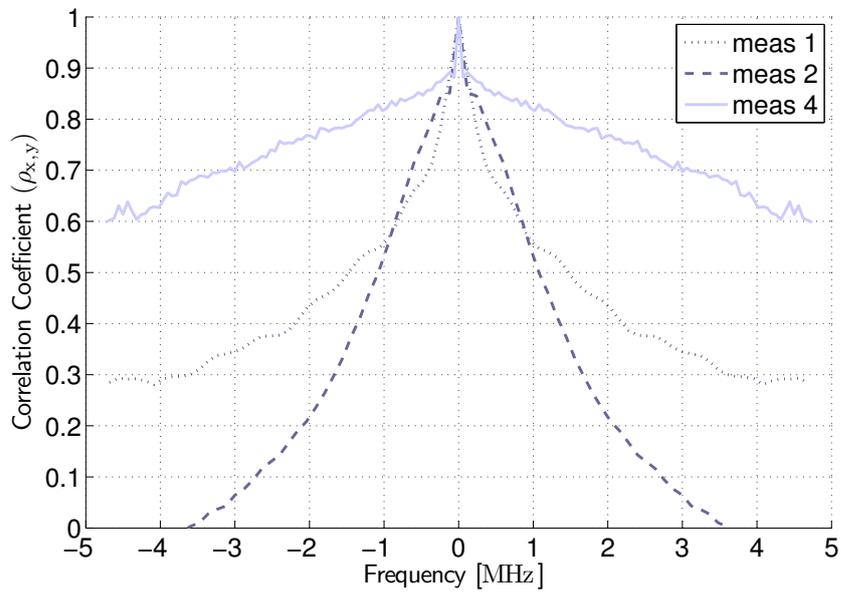


Figure 5.4: Comparison of the frequency diversity in the different measurements. The plot shows the frequency diversity of user 1 for all three measurements.

For all 4 users, the correlation coefficient decreases as the frequency increases. This is what we would expect: the further two sub-channels are separated in frequency, the less correlated they should be. Looking at the zoom in the top left corner of Figure 5.3, we can see that the coherence bandwidth varies between $b_{\text{coh}} = 115$ kHz (user 2) and $b_{\text{coh}} = 790$ kHz (user 4). Those values match the analytical values using Equation (2.3) and the standard deviation of the delay spread for a suburban environment from Table 2.1, for which the analytically calculated coherence bandwidth is between $b_{\text{coh}} = 80$ kHz and $b_{\text{coh}} = 800$ kHz.

In Figure 5.4 we plot the correlation coefficient of user 1 for all three measurements. Looking again at the 0.9 correlation coefficient, the coherence bandwidth shows similar values for measurements 2 and 4 as for measurement 1. However, the slope of the correlation coefficient for measurement 4 is not as steep as for the other two measurements. Remember that measurement 4 was taken in a scenario with LOS between transmitter and receiver. In such a scenario a lower standard deviation of the delay spread is to be expected, which explains the difference between measurement 4 and the other two measurements.

Figures 5.5 and 5.6 show the measurement results for the time domain. From Figure 5.5, the coherence time for measurement 1 can be estimated. Looking at the zoom in the top left corner shows that the coherence time for this measurement is between $t_{\text{coh}} = 1.5$ ms and $t_{\text{coh}} = 2$ ms. According to Equations (2.1) and (2.2) a coherence time of 2 ms corresponds to an average velocity of $v = 50$ km/h which means that the measurement results match the analytical values.

In Figure 5.6 we again compare the measurement results for user 1 in all three measurements. The figure shows that the coherence time is significantly larger for measurement 2 ($t_{\text{coh}} = 20$ ms) and 4 ($t_{\text{coh}} = 15$ ms), which corresponds to an average velocity of about $v = 5$ km/h. Thus, also for the coherence time the measurement results confirm the analytical calculations.

We also investigated the spatial diversity and calculated the correlation coefficients between the measurements taken by the individual users. We found the correlation coefficient to be very low (below 0.15) for all three measurement campaigns, which indicates that there is no spatial correlation between the measurements taken by the individual users. Those results confirm our considerations from Section 2.3.1, namely that measurements are already uncorrelated once the antennas are only a few wavelengths apart, such that in the case of multiple sensors, measurements can always be assumed to be i.i.d. and, thus, to contribute to the spatial diversity of the measurement results.

5.3.2 Diversity Limits

We have shown earlier that there are strict limits on the diversity achievable in the time and frequency domain: it is limited by the PU bandwidth (b_{pu}) and the maximum interference time (t_{max}) on the one hand and by the coherence bandwidth (b_{coh}) and coherence time (t_{coh}) on the other hand. For spatial diversity there is — in theory — no such limit. But the time, spectrum, and processing power required to exchange and merge the sensing results from multiple sensors put practical limits on the achievable

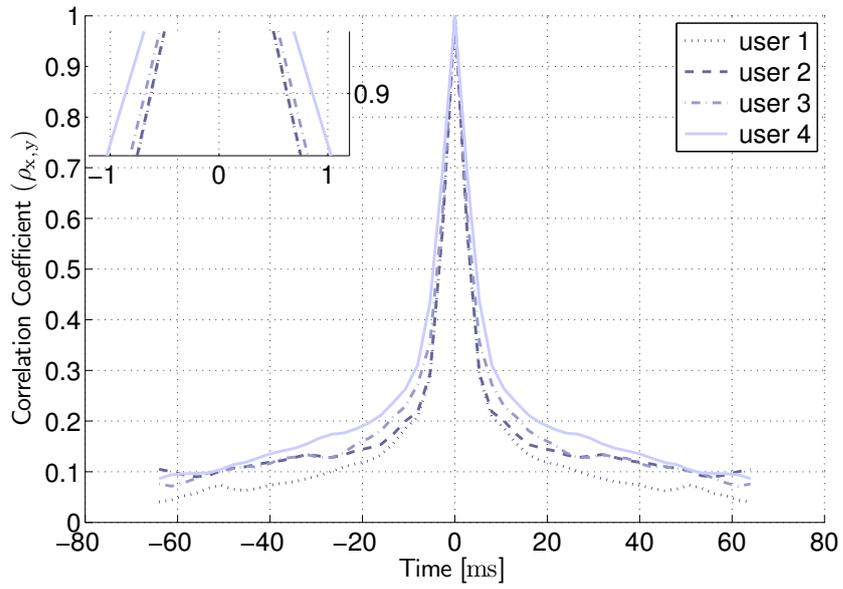


Figure 5.5: Time diversity in measurement 1

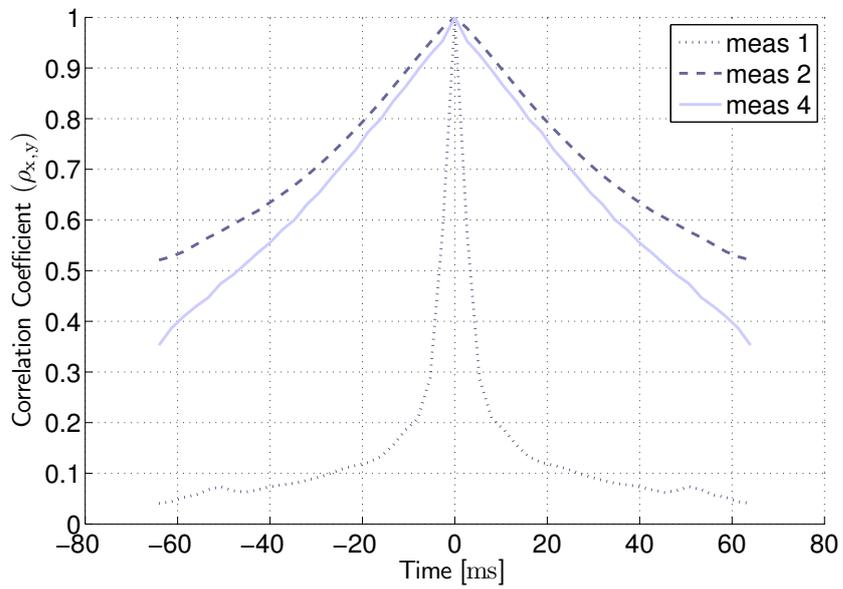


Figure 5.6: Comparison of the time diversity in the different measurements. The plot shows the time diversity of user 1 for all three measurements.

spatial diversity. However, we do not investigate these practical limits of spatial diversity here and focus instead on the diversity limits of the local sensing process, i.e., time and frequency diversity.

Obviously, the diversity limits of the local spectrum sensing process depend on the details of the PU and on the scenario the CR system is operating in. Thus, it is not possible to make general statements on the limits of achievable diversity. Instead, we look at selected scenarios and try to give a general idea of the parameters to consider when using diversity for spectrum sensing.

Based on the assumption that sensing samples are i.i.d., if taken in time intervals not smaller than the coherence time and frequency intervals not smaller than the coherence bandwidth, it does not matter whether to use time or frequency diversity for the sensing process. Both diversity approaches lead to the same improvement of the sensing result. The t_{coh} and b_{coh} constraints are also the first limit of the achievable diversity. They specify the minimum resolution at which sensing samples can be taken. Using this minimum resolution, the number of sensing samples, and thus the achievable diversity, is limited by b_{pu} and t_{max} . The maximum diversity that can be achieved in the time domain, thus, depends on the agility of the PU (which influences t_{max}) and on the carrier frequency of the PU bands and the average velocity (which influence t_{coh}) and computes to $t_{\text{coh}}/t_{\text{max}}$. For the frequency domain, the maximum achievable diversity depends on the PU bandwidth (b_{pu}) and on the environment the PU system is operating in (which influences b_{coh}) and computes to $b_{\text{coh}}/b_{\text{pu}}$. To be able to compare the maximum diversity achievable in both domains, we define the diversity difference fraction (d_{d}) as the difference between the maximum time and frequency diversity divided by the maximum total diversity:

$$d_{\text{d}} = \frac{\frac{t_{\text{max}}}{t_{\text{coh}}} - \frac{b_{\text{pu}}}{b_{\text{coh}}}}{\frac{t_{\text{max}}}{t_{\text{coh}}} + \frac{b_{\text{pu}}}{b_{\text{coh}}}} \quad (5.16)$$

If $d_{\text{d}} = 0$, the achievable time and frequency diversity is equal, for $d_{\text{d}} > 0$ it specifies how much more time diversity is achievable, for $d_{\text{d}} < 0$ how much more frequency diversity is achievable.

Diversity Tradeoff in the TV Bands

As an example, we consider the diversity tradeoff for the TV bands. Based on the U.S. specifications, we consider TV channels with a bandwidth of $b_{\text{pu}} = 6$ MHz as PU bands, which operate in the frequency range around $f_{\text{c}} = 600$ MHz. We set the maximum interference time to $t_{\text{max}} = 2$ s, which corresponds to the specifications used within the IEEE 802.22 standardization [82].

For the TV bands an urban or suburban scenario is most applicable, which, based on Equation (2.3) and Table 2.1, results in a coherence bandwidth between $b_{\text{coh}} = 6.4$ kHz and $b_{\text{coh}} = 800$ kHz. The TV transmitters are static; however, the CR network might be mobile. Thus, assuming the average velocity varying between $v = 0.5$ m/s (slow walking speed) and $v = 15$ m/s (city speed limit for cars), the coherence time is, according to Equations (2.1) and (2.2), between $t_{\text{coh}} = 7.5$ ms and $t_{\text{coh}} = 225$ ms.

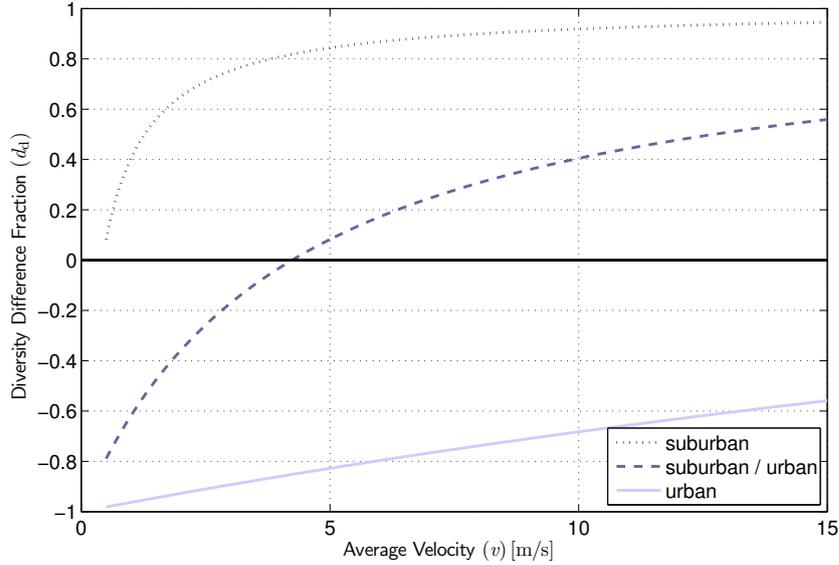


Figure 5.7: Diversity tradeoff for TV bands

We investigate the diversity difference fraction (d_d) for this example in Figure 5.7. It shows d_d against the average velocity (v) of the CR system for three different scenarios (suburban $\sigma_{ds} = 0.2 \mu s$, suburban / urban $\sigma_{ds} = 2 \mu s$, and urban $\sigma_{ds} = 25 \mu s$). The figure shows that the diversity tradeoff depends heavily on the environment and the average velocity. Generally speaking, a more urban environment results in a higher standard deviation of the delay spread and thus a smaller coherence bandwidth, which increases the achievable diversity in the frequency domain. On the other hand, a higher average velocity results in a smaller coherence time and thus in an increase of the achievable time diversity. In suburban environments, the diversity achievable in the time domain is almost always higher than that achievable in the frequency domain. For suburban environments with a very low standard deviation of the delay spread (dark dotted line) and normal walking speed (and higher velocities) the time diversity is at least 80% of the total diversity achievable. Only for the urban environment with a very high standard deviation of the delay spread (light solid line) the achievable frequency diversity is always bigger than the time diversity.

Diversity Tradeoff in the Cellular Bands

As examples for the cellular bands, we investigate the diversity tradeoff for the UMTS and LTE bands. The UMTS bands have a bandwidth of $b_{pu} = 5$ MHz and the LTE bands a bandwidth of (up to) $b_{pu} = 20$ MHz. For both we assume a carrier frequency of $f_c = 1.9$ GHz, which corresponds to the allocation in Germany. Since secondary usage in the cellular bands is not considered by regulation so far, there are no official values for t_{max} available. We choose $t_{max} = 0.5$ s which we believe should be an acceptable value.

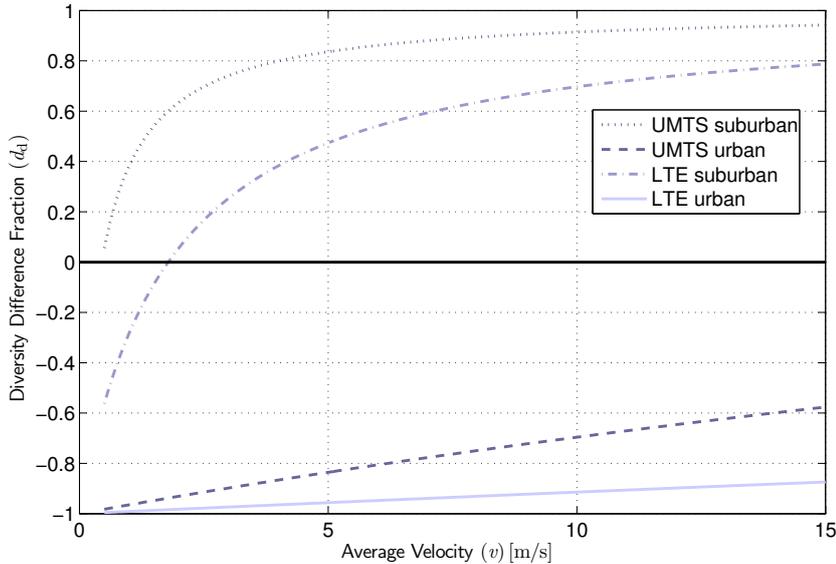


Figure 5.8: Diversity tradeoff for cellular bands

The scenario for the cellular bands is pretty similar to the scenario used for the TV bands above. We consider a coherence bandwidth based on a suburban / urban scenario ($b_{\text{coh}} = 6.4 \text{ kHz}$ to $b_{\text{coh}} = 800 \text{ kHz}$). Furthermore, the primary base stations are fixed while the primary terminals and the CR network might be mobile, so we consider an average velocity between $v = 0.5 \text{ m/s}$ (slow walking speed) and $v = 15 \text{ m/s}$ (city speed limit for cars), resulting in a coherence time between $t_{\text{coh}} = 2.4 \text{ ms}$ and $t_{\text{coh}} = 71 \text{ ms}$.

In Figure 5.8 we show the diversity tradeoff for the cellular bands. We only show graphs for the two extreme environments (urban, very high standard deviation of the delay spread and suburban, low standard deviation of the delay spread). The general observations are the same as for the TV bands, namely that in suburban environments the time diversity dominates, whereas in urban environments the frequency diversity dominates. Comparing the graphs for UMTS and LTE we can see the influence of the PU bandwidth. The higher b_{pu} of LTE leads to a general increase of the frequency diversity.

5.4 Related Work

There has been a lot of research activity on analytically investigating the influence of the local sensing time on the reliability and quality of the sensing process. For an overview, refer to [176] and references therein. The influence of the sensing bandwidth used has received far less attention in the research community so far. Mostly, it is assumed that the whole PU bandwidth is used for the sensing process.

Note that although the influence of the local sensing time has been extensively investigated, there are only very few investigations on the exploitation of time *diversity*.

For the investigation of frequency diversity we are not aware of a single publication. The impact of fading on energy detection has been investigated in [134, 145]; however, the authors did not investigate the potential gains from exploiting the diversity of the received signal. In [40, 41] the diversity gain in energy detection of signals over various fading channels is investigated using equal gain combining, selection combining, and switch and stay combining. The authors in [68] focus on Nakagami fading channels and equal gain combining and the authors of [10] investigate the diversity gain in energy detection considering multi-path fading and shadowing. While all investigations show that exploiting diversity improves the detection accuracy, they all use the information theoretic concept of “diversity branches”. Unfortunately, this concept cannot be easily mapped to time or frequency diversity as investigated in this thesis.

In [99] the diversity gain dependent on the time diversity is investigated using energy detection. The results of [99] are used in this thesis and generalized to the case of time and frequency diversity.

The major research focus related to diversity approaches for spectrum sensing has been on spatial diversity and we only list some exemplary work here. For distributed sensing, there is research on the area within which CRs should cooperate [136] and a lot of research on which sensing results to transmit and in which form (raw sensing data vs. one-bit decisions) [120, 130, 144]. An overview of possible fusion techniques for sensing data can be found in [26]. The quality improvement of the sensing process by simply “or”ing the sensing results is presented in [57, 156]. In [149] such simple approaches are compared with a more sophisticated approach using partial statistical knowledge.

5.5 Conclusions

In this chapter we have investigated whether and how **reliable protection of the PU communication**, quantified by the probability of false negatives (P_{fn}), is possible with a spectrum sensing based detection process. Specifically, we have focused on simple, energy detection-based spectrum sensing being a cheap and simple sensing approach. While reliable (low P_{fn}) and high quality (low probability of false positives (P_{fp})) spectrum sensing is possible for the energy detector from an information-theoretic point of view, there are certain practical limitations. The most important limitation is the diversity required for the sensing process, which is really high. We have shown that reliable PU protection is also possible using very limited diversity for the spectrum sensing process. Dropping the requirement of a low probability of false positives (which has no influence on the reliable PU protection), a very low probability of false negatives can also be achieved using very limited diversity.

We have investigated the diversity in the time, frequency, and spatial domain. Evaluating measurements done in the LTE bands we have shown that the analytical models can be applied to estimate diversity. Since time and frequency diversity are strictly limited by the coherence time and coherence bandwidth on the one hand, and by the maximum interference time and PU bandwidth on the other hand, we have investigated the maximally achievable diversity for selected scenarios. From our investigations, we

can conclude that the higher the average velocity (of PU and CR system), the more dominating is the time diversity. However, except for urban environments with a very high standard deviation of the delay spread, the achievable time diversity is usually bigger than the frequency diversity also for low average velocity.

The two main conclusions for the CR system design presented in Chapter 6 are:

1. Reliable PU protection is possible using simple energy detection. For the case of low diversity in the measurement results, however, the probability of false positives is really high.
2. Generally speaking, the achievable diversity is much higher in the time domain compared to the frequency domain. The achievable frequency diversity is only dominating for urban environments with a very high standard deviation of the delay spread.

CHAPTER 6

Cognitive Radio System Design

In this chapter the CR system design used in this thesis is proposed. The general concept for the design has been published in [17, 20]. The proposed design features a flexible periodic sensing process first described in [164, 165] whose diversity can be dynamically adapted to the given requirements. Secondary User Links (SULs) used for secondary communication are created in a way which is robust to reconfigurations due to the detection of PUs. Furthermore, mechanisms to compensate for reconfigurations are proposed based on work presented in [164, 165, 167]. The chapter concludes with a feasibility assessment for the realizability of the proposed design.

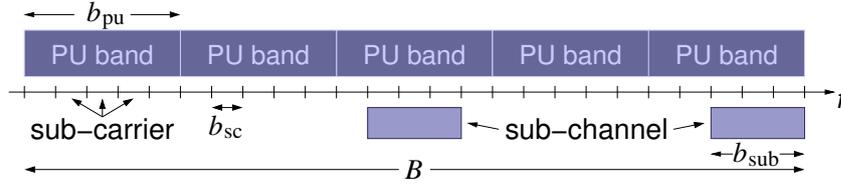


Figure 6.1: Cognitive Radio system bandwidths

The proposed system design is based on CORVUS [17, 20] with an underlying OFDM system. The CR system covers a bandwidth of B Hz, divided into N_{ofdm} OFDM sub-carriers with a fixed bandwidth of $b_{sc} = B/N_{\text{ofdm}}$. There are N_{pu} non-overlapping PU bands of bandwidth b_{pu} within the CR system bandwidth B . The relations between the different bandwidths used to describe our system design are shown in Figure 6.1.

6.1 Secondary User Links

Non-contiguous SULs, as proposed in the original CORVUS design [17], are used for data transmission within a Secondary User Group (SUG). A sub-channel, consisting of N_{sc} OFDM sub-carriers, is the basic allocation unit for secondary communication. A SUL consists of N_{sub} sub-channels having equal sub-channel bandwidth (b_{sub}), which are distributed over several PU bands with the constraint of maximally one sub-channel per PU band. The resulting SUL bandwidth (b_{sul}) computes to $b_{sul} = N_{sub} \cdot b_{sub} = N_{sub} \cdot N_{sc} \cdot b_{sc}$. Figure 6.1 shows a SUL with $N_{sub} = 2$ sub-channels, each consisting of $N_{sc} = 3$ sub-carriers, thus, having a bandwidth of $b_{sul} = 2 \cdot b_{sub} = 6 \cdot b_{sc}$.

The proposed SUL design already provides some robustness against secondary QoS degradation due to link reconfigurations. If a PU appears in a PU band used by the SUL, only one sub-channel of the SUL needs to be replaced — the rest of the SUL is not affected. This implies that in the proposed CR system sub-channels can be dynamically removed and added to the SUL without the need to interrupt the usage of the other sub-channels of the SUL.

The concept of non-contiguous SULs offers a great flexibility in the creation and use of SULs. A SUL of the same SUL bandwidth (b_{sul}) can be created in many different ways by varying the number of OFDM sub-carriers per sub-channel (N_{sc}) and with this the number of sub-channels of a SUL (N_{sub}). Thus, the scattering of a SUL over different PU bands and with this the sensitivity to the appearance of a single PU can be varied. This concept is illustrated in Figure 6.2, which shows three different SULs all having the same SUL bandwidth of $b_{sul} = 4$ sub-carriers. Whereas SUL 1 is created with only $N_{sub} = 1$ sub-channel consisting of $N_{sc} = 4$ sub-carriers, and, thus, is completely placed within one PU band, SUL 3 is scattered over 4 PU bands by dividing it into $N_{sub} = 4$ sub-channels consisting of $N_{sc} = 1$ sub-carrier each. Note that, for the presented example, the instantaneous link capacity would drop to zero for SUL 1 if the respective PU appears, whereas SUL 3 would still be able to maintain 3/4 of its link capacity.

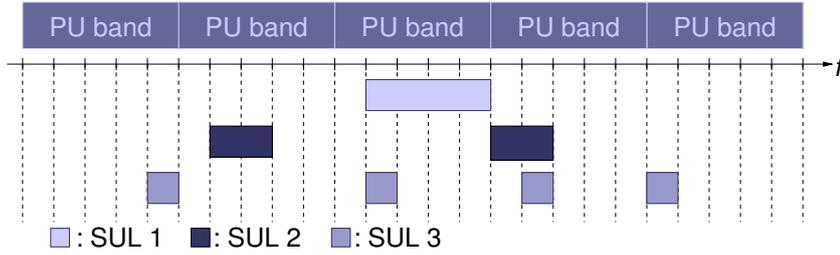


Figure 6.2: Example distributions of sub-channels for a SUL with bandwidth $b_{\text{sul}} = 4$ sub-carriers.

The above definition of SULs also allows a great flexibility for the offered maximum link capacity. By adding or removing sub-channels from the SUL the SUL bandwidth can be dynamically adapted to the requirements of the applications.

6.2 Periodic Spectrum Sensing

The periodic spectrum sensing process consists of a local sensing phase, executed locally within each CR and a distributed sensing phase where the local sensing samples are merged to exploit the spatial diversity of the individual sensing results. The periodic sensing approach used for our CR system should be designed to support secondary QoS maintenance, which is — especially for the local sensing phase — not straight forward, as shown below.

6.2.1 Local Sensing

During the local periodic sensing process, the spectral resources used for spectrum sensing cannot be used for data transmission at the same time. Furthermore, to assure the reliability of the sensing process, also neighboring CRs should not use the respective spectrum during the sensing process as elaborated in Section 5.2.

Interrupted Sending

The most straight forward approach to organize the periodic sensing process and to adhere to the above constraints is interrupted sending as shown in Figure 6.3(a). Interrupted sending is also the approach most frequently followed in the research community so far. It is also applied within the IEEE 802.22 standard [32, 82]. As the name suggests, secondary data transmission has to be periodically interrupted to perform sensing. Note that, in general, interruptions of individual SULs do not need to be synchronized. If the sensing process detects a PU in an opportunistically used PU band, the SUL has to vacate it and — usually — shifts the SUL to another, unused band.

The inherent problem of the interrupted sending approach is that usage of the SUL has to be periodically interrupted. Depending on the time required for the sensing process, these interruptions can severely degrade the secondary QoS. Several studies

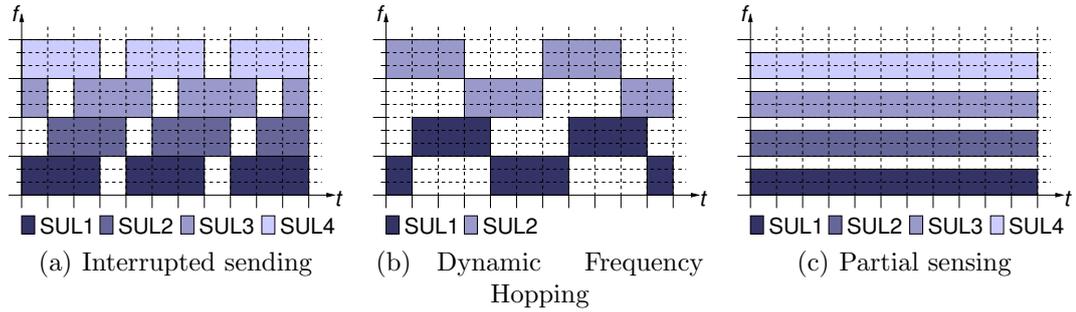


Figure 6.3: Approaches for periodic sensing organization. The colored areas represent spectrum used for different communications links (SULs), the white areas spectrum available for sensing.

about the sensing time for detecting TV signals indicate that tens or even up to hundreds of milliseconds are required for the sensing process [25, 52, 138]. Thus, the interrupted sending approach is not appropriate for a CR design which supports secondary QoS.

Dynamic Frequency Hopping

To overcome the periodic interruptions, we have proposed Dynamic Frequency Hopping (DFH) [71]. We also proposed DFH as an alternative to interrupted sending for the IEEE 802.22 standard [31]. The basic idea of DFH is shown in Figure 6.3(b). It follows the idea of performing sensing and data transmission in parallel in different PU bands and switching them cyclically in such a way that — with exception of the very short switching gap — continuous sending can be assured. One drawback of DFH is that a more complex radio front end — being able to sense and send in parallel — is needed.

Since DFH avoids the periodic interruptions of the usage of the SUL, it is a potential candidate for a secondary QoS aware CR system. However, DFH is designed to work with contiguous SULs. Although it could theoretically also be applied to non-contiguous SULs, the periodic switching of the (potentially many) sub-channels would result in a very high complexity and a high coordination overhead. Nevertheless, some performance results in comparison to the partial sensing approach introduced below are shown in Chapter 7 and Chapter 8 also for the DFH approach.

Partial Sensing

Based on the CORVUS system design [17, 20] and with non-contiguous SULs in mind we have developed the partial sensing approach [165], which is shown in Figure 6.3(c). The idea is that some part of the PU band is always left idle to perform sensing. The rest of the PU band can be continuously used for the SUL in parallel. If a PU is detected in its band, data transmission in the affected PU band has to be discontinued and the SUL has to be shifted to another, unused band. Similar to the DFH approach, the radio front end has to be capable of performing sending and sensing in parallel.

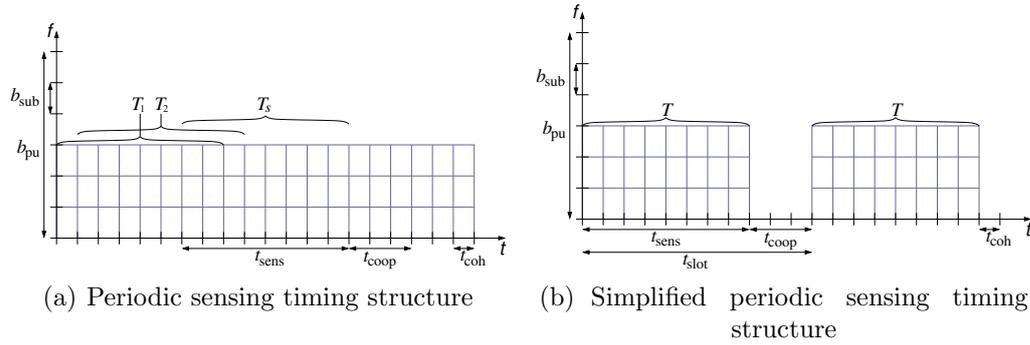


Figure 6.4: Time domain sensing model

Partial sensing avoids periodic interruptions of the data transmission on the one hand, and on the other hand supports the use of non-contiguous SULs and is, thus, the ideal candidate for our CR system design. We assume that the sensing bandwidth (b_{sens}) is flexible and can be adapted to the requirements and constraints of the CR system. Without loss of generality, b_{sens} is assumed to be an integer multiple of the coherence bandwidth (b_{coh}), resulting in $M_b = b_{\text{sens}}/b_{\text{coh}}$ i.i.d. sensing samples recorded in the frequency domain. The diversity gain in the frequency domain (as investigated in Section 5.3) can be dynamically adapted by varying the number of sensing samples in the frequency domain (M_b).

6.2.2 Distributed Sensing

In addition to the local sensing performed in each CR, the SUG members participate in a distributed sensing process. The number of cooperating sensors (m) can be varied and adapted to the requirements of the sensing process. The specific selection criteria for which CRs should participate in the distributed sensing process are beyond the scope of this thesis. For the proposed CR system, we assume a distributed sensing process based on hard decision combining of one-bit sensing results of the individual CRs. The incentive for this is the low complexity of achieving a joint decision and the very low required signaling overhead compared to soft decision combining.

The details of the protocols for this distributed sensing process are beyond the scope of this thesis. However, distributed sensing takes time to exchange the local sensing results and make a decision. This required time is denoted by t_{coop} in this thesis.

6.2.3 Time Domain Model

To assure the protection of PUs, it is crucial that a PU is detected within t_{max} after its appearance. Thus, the *whole* sensing process (local plus distributed sensing) is strictly upper bounded by t_{max} ($t_{\text{sens}} + t_{\text{coop}} \leq t_{\text{max}}$).

To better understand the details of the time structure for the periodic sensing process, let us conceptually divide the time axis into slots having the length of a coherence

time (t_{coh}). As shown above, the sensing process can then record $M_b = b_{\text{sens}}/b_{\text{coh}}$ independent sensing samples within each slot. For simplicity, let us further assume that a PU can only appear at the beginning of such a time slot. For the local sensing process to be able to reliably detect PUs no matter in which slot they appeared, spectrum sensing has to be done continuously. This is shown in Figure 6.4(a), where each rectangle corresponds to one mini-block of the size $b_{\text{coh}} \cdot t_{\text{coh}}$ used to produce one sensing sample and $M_b = 3$ sensing samples are recorded in parallel in one time slot.

Assuming the energy detector as described in Section 5.2.2 and assuming a local sensing time (t_{sens}) of $M_t = 8$ slots as shown in Figure 6.4(a), the test statistic defined in Equation (5.11) at the end of slot s (T_s) is computed as a moving average using the $M = M_t \cdot M_b$ sensing samples of the last M_t time slots. At the end of each time slot, T_s is tested against the detection threshold, and if it exceeds the threshold, i.e., a PU is detected by the local sensing process, the distributed sensing process is triggered, which takes 3 time slots in the example of Figure 6.4(a).

Without loss of generality a simplified model is assumed in the remainder of this thesis to enable a more descriptive presentation of the general concepts. Time is slotted into slots of length t_{slot} with $t_{\text{slot}} \geq t_{\text{sens}} + t_{\text{coop}}$. A PU can only appear at the beginning of a slot and, thus, there is also only one sensing process per slot, starting at the beginning of the slot (no moving average). After the local sensing process is finished, the collection of sensing samples is interrupted and the distributed sensing process is started, if a PU was detected. At the beginning of the next slot, the local sensing process starts to collect sensing samples again. The concept of this simplified model is shown in Figure 6.4(b). Within the assumed model t_{slot} is, thus, also the time a PU band is used for data transmission before it is revalidated by the sensing process, which implies $t_{\text{slot}} \leq t_{\text{max}}$.

6.3 Link Reconfiguration

The link reconfiguration process is responsible for adjusting the sub-channels used by a SUL to the current interference environment. Specifically, sub-channels on which a PU was detected have to be excluded from the SUL and new sub-channels have to be added to compensate for the lost ones. For the proposed system design, it is assumed that sub-channels can be seamlessly removed from and added to a SUL without affecting the data transmission on the remaining sub-channels of the SUL.

Excluding of sub-channels from the SUL is very time critical. To ensure the reliable protection of the PU, a sub-channel on which a PU appeared has to be vacated within the maximum interference time (t_{max}). Adding of new sub-channels in contrast is not as time critical, since no harm is done to PUs. However, until new sub-channels are added to the SUL, the capacity of the SUL is reduced, which impacts the secondary QoS.

6.3.1 Removing Interfered Sub-Channels

Due to the close connection between distributed sensing and the excluding of sub-channels, these two steps are often combined. In the distributed sensing process, all

SUG members need to achieve a joint view on which sub-channels are used by PUs, and in the excluding of sub-channels process, all SUG members have to be informed about which sub-channels to exclude from the SUL.

For the proposed CR design, we assume such a combination of the two steps. At the end of the periodic sensing process, all members of the SUG know on which of the sub-channels used by the SUG a PU was detected. They can, thus, immediately exclude the corresponding sub-channels from the SUL. Due to the time constraints of the periodic sensing process ($t_{\text{sens}} + t_{\text{coop}} \leq t_{\text{max}}$), removal of interfered sub-channels is achieved within t_{max} , i.e., the PU protection requirement is fulfilled in this case.

The above statement implies a fundamental assumption. All SUG members *correctly* receive the periodic sensing results, i.e., control communication is 100% reliable. This is a very hard assumption for wireless communication in general, and even more for secondary communication based on Dynamic Spectrum Access (DSA).

Without going into the detailed protocol implementations, which are beyond the scope of this thesis, we proposed a soft-state-based approach [86] for the reliable and timely vacation of sub-channels on which a PU was detected. The basic idea is that each PU band detected to be available for secondary communication has an associated lifetime. This lifetime is upper bounded by t_{max} , and sub-channels out of the respective PU band can only be used for a SUL until the lifetime expires. Every time the periodic sensing process revalidates the availability of a PU band, i.e., the distributed sensing results are received correctly, the lifetime is renewed. If, however, no such update is received, sub-channels from the respective PU band cannot be used anymore. Thus, the periodic sensing process should be dimensioned such that there is a high probability that the updated periodic sensing results are available before the lifetime expires.

The above approach can be easily adapted to the quality of the control channel and the requirements of the link reconfiguration process, by adjusting the frequency of the periodic sensing process. E.g., by executing the periodic sensing twice as often as required ($t_{\text{sens}} + t_{\text{coop}} = 0.5 \cdot t_{\text{max}}$) and having a maximum lifetime of t_{max} , the probability that the lifetime of a PU band expires due to packet loss is reduced by 50%.

6.3.2 Adding New Sub-Channels

To compensate for the excluded sub-channels and maintain the secondary QoS, new sub-channels need to be added to the SUL. However, before adding new sub-channels to the SUL, it has to be ensured that the sub-channels in question are currently not used by the respective PU, i.e., that their lifetime is not expired. Furthermore, the members of a SUG have to *agree* on which sub-channels out of the available ones to add to the SUL.

There are two general approaches to updating the lifetime of sub-channels intended to be added to a SUL, a *proactive* and a *reactive* approach. In the proactive approach, the periodic sensing process not only senses the PU bands out of which sub-channels are currently used for the SUL, but senses the whole spectral operation range of the CR or at least an additional subset of PU bands out of the PU bands currently not used. Using this approach, each SUG member already knows at the end of the periodic sensing process

which sub-channels are available to be added to the SUL. In the reactive approach, the periodic sensing process only senses PU bands used by the SUL. Consequently, the reconfiguration process has to trigger a new sensing process on a set of PU bands currently not used by the SUL to find new sub-channels to be added.

Note that in both approaches the SUG members have to have a common understanding on which additional PU bands to sense (if the whole operation range cannot be sensed simultaneously). For this additional policies and protocols are required.

Once the sensing results are available, the SUG members have to negotiate which (and possibly also how many) sub-channels to use out of the available ones. Whereas in most cases (at least) the lost sub-channels should be replaced by new ones, it might also be the case that the number of sub-channels used for the SUL should be increased or decreased. This might be necessary, e.g., to adjust the capacity of the SUL or the redundancy used for the SUL (explained in the next section).

For the proposed system design we assume a proactive reconfiguration approach. Since an OFDM-based underlying system is assumed, spectrum sensing can be performed within in the whole operation range in parallel. Note however, that the results presented in this thesis are also applicable to reactive approaches.

The details of the required protocols and algorithms are beyond the scope of this thesis and subject to future work. Link reconfiguration is modeled by the link reconfiguration time (t_{reconf}), being the time from removing the interfered sub-channels from the SUL until the new sub-channels are selected and added to the SUL.

6.4 Secondary QoS Maintenance

The CR system presented above is already designed to reduce the variation of the link capacity offered by a SUL. Using a partial sensing approach eliminates the need to interrupt usage of the SUL sub-channels for spectrum sensing and the non-contiguous SUL approach reduces the temporal capacity loss in case a PU was detected to one sub-channel. However, as explained above, the reconfiguration of a SUL still requires time during which the link capacity of the SUL is reduced. For the presented CR system design, two different approaches are proposed to compensate the temporal degradation of the SUL link capacity due to reconfigurations: a *redundancy* and a *resource reservation* approach.

6.4.1 Redundancy Approach

The redundancy approach follows the commonly applied compensation approach of over-provisioning. Redundancy is added to the SUL such that even if data transmission on *some* part of the SUL has to be interrupted, there is still enough spectrum available for the SUL to offer the minimum bandwidth.

The basic idea is to add X redundant sub-channels to the N_{sub} sub-channels of the SUL and apply coding in such a way that the receiver can decode the message, if *any* N_{sub} out of the $N_{\text{sub}} + X$ sub-channels are received. Using this approach the SUL can

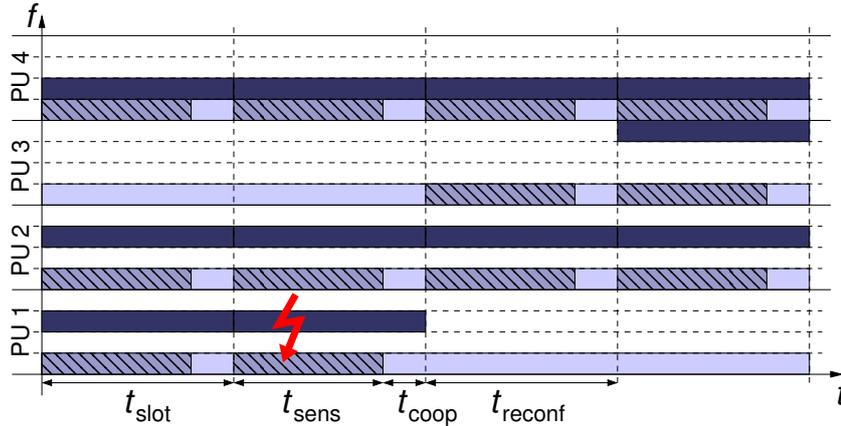


Figure 6.5: Redundancy example

tolerate the concurrent appearance of PUs on up to X sub-channels and still maintain the required link capacity. More details on this coding technique applied in the context of Cognitive Radio can be found in [167].

Figure 6.5 shows a conceptual example of the assumed redundancy approach. In the example, the SUG operates in a spectrum range covering four, equally sized PU bands. Each PU band consists of four sub-channels and one sub-channel is reserved for the sensing process (the first, light blue sub-channel of each PU band in the figure). The SUG uses a SUL with a bandwidth of $N_{\text{sub}} = 2$ sub-channels and, to maintain the secondary QoS, $X = 1$ redundant sub-channel is added resulting in a SUL consisting of $N_{\text{sub}} + X = 3$ sub-channels (the dark blue sub-channels in the figure). The shaded areas indicate the time used for the local sensing process, which cannot span the whole slot, since additional time for the cooperative sensing process is needed, as explained above.

Assuming that PU 1 appears in the second slot as shown in Figure 6.5, the local and cooperative sensing process assures that all SUG members know of the appearance at the end of the slot and can discontinue the use of the respective sub-channel. Since $t_{\text{sens}} + t_{\text{coop}} \leq t_{\text{slot}} \leq t_{\text{max}}$, the PU protection constraint is satisfied. In the next slot, the capacity of the SUL is reduced to two sub-channels; however, due to the added redundancy, the secondary QoS can still be provided. The SUG has a whole slot to find another sub-channel to be added to the SUL ($t_{\text{reconf}} \leq t_{\text{slot}}$), such that in the following slot the full redundancy is reestablished for the SUL.

6.4.2 Resource Reservation Approach

An alternative to the redundancy approach is a *resource reservation* approach in which some backup sub-channels are reserved for the SUL, which are instantly available if the link needs to be reconfigured. This means that the SUG has to maintain a set of backup sub-channels, which have to be regularly sensed (within the periodic sensing process) to assure their immediate availability. The SUL can then be seamlessly switched to use the

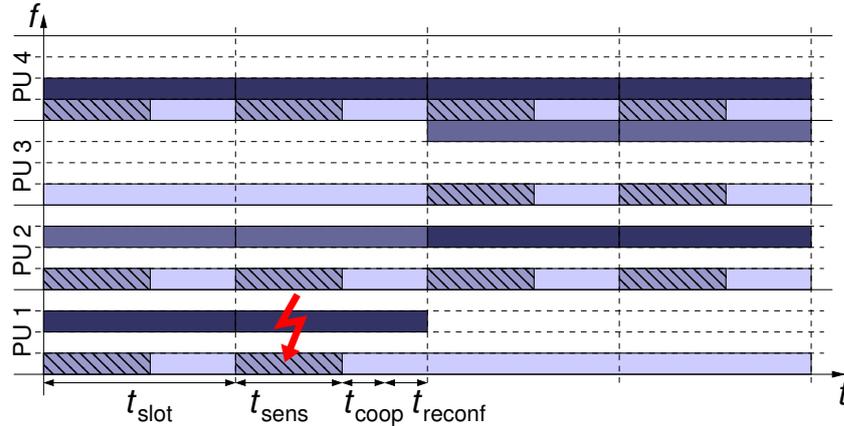


Figure 6.6: Resource reservation example

backup sub-channels.

Figure 6.6 shows the same scenario as Figure 6.5 but this time using the resource reservation approach. To compensate for reconfiguration losses, $X = 1$ sub-channel is reserved for the SUL (the sub-channel in PU 2 indicated by the slightly brighter color than the color of used sub-channels in the figure).

Assuming that PU 1 appears in the second slot as shown in Figure 6.6, the periodic sensing process will (hopefully) detect the appearance and notify all SUG members. Within the reconfiguration period (t_{reconf}) a replacement for the interfered sub-channel has to be selected from the backup sub-channels¹ and in addition a new backup sub-channel has to be selected to refill the pool of backup spectrum. Thus, at the end of the slot a new sub-channel is selected such that the SUL can be seamlessly switched to use the new sub-channel as shown in Figure 6.6. Since for the resource reservation approach $t_{\text{sens}} + t_{\text{coop}} + t_{\text{reconf}} \leq t_{\text{slot}} \leq t_{\text{max}}$, the PU protection requirement is satisfied.

6.4.3 Comparison and Spectral Overhead

One difference of the resource reservation approach compared to the redundancy approach is that the reconfiguration period is much more time critical. Whereas reconfiguration is only bounded by the length of a slot ($t_{\text{reconf}} \leq t_{\text{slot}}$) in the redundancy approach, it has a direct influence on the time available for the sensing process in the resource-reservation approach. In practice, this means that less time diversity can be used in the local sensing process.

The spectral overhead of the redundancy and resource reservation approach is determined by the number of redundant / reserved sub-channels of a SUL (X). The more redundant / reserved sub-channels are added to the SUL, the bigger is the spectral overhead. More spectral overhead, on the other hand, also results into a better secondary

¹Although only one backup sub-channel is shown in the example, there will generally be more than one, so that a selection process has to take place if the SUL needs to be reconfigured.

QoS support. This tradeoff is evaluated in Chapter 8.

Note that, considering multiple SULs in the same area, the spectral overhead in case of the resource reservation approach could be potentially reduced by exploiting the multiplexing gain if those SULs *jointly* maintain a set of backup sub-channels. Since the redundant sub-channels are actually used in the redundancy approach such a multiplexing is not feasible for this approach.

However, note that joint maintenance would require different SUGs to coordinate their backup sub-channels requiring communication overhead which would increase the spectral overhead again. Since in the presented work only one single SUL is considered, the mechanisms and protocols required for management of backup sub-channels and potential backup sub-channel sharing are subject to future work.

6.5 Feasibility Assessment

In this section we highlight some aspects concerning the realizability of the proposed CR system design. Within this thesis, the proposed design has not been implemented and tested in hardware so we want to at least assess the feasibility of the most important aspects of our design.

The most fundamental limitation of the proposed CR design is the availability of low cost, wide-band OFDM front-ends. Ultimately, our system should be able to operate in a wide spectrum range covering several hundred megahertz of spectrum licensed to a large number of diverse PUs. Wireless OFDM-based systems today operate on much lower bandwidth. For WiMAX and LTE, OFDM front-ends with 20 MHz of bandwidth are available and for LTE-Advanced front-ends with up to 40 MHz are expected. However, within the research community prototype platforms are already available covering a much higher bandwidths. The newest version of the USRP SDR platform (N200 series) [45], which is widely used in the CR research community, supports the processing of signals of up to 100 MHz of bandwidth. Within the GENI project [55] currently a front-end supporting even up to 500 MHz is developed.

Although our proposed CR design is targeted for using a wide spectrum range, for which low cost front-ends will most likely not be available in the near future, we still believe that the proposed design has great potential. First of all, recent developments show the rapid progress on the semiconductor market and the availability of research prototypes indicates that such wide-band front-ends will be eventually available at low cost. Furthermore, the proposed design is not limited to such wide-band usage. Principally, it can also be applied to rather small operation ranges as long as it covers at least some different PU bands. Due to regulatory constraints, early deployments will in fact most likely be limited to relatively small spectrum bands licensed to homogeneous PUs. Considering, e.g., the 2G cellular spectrum in the 900 MHz band, there are 35 MHz of spectrum divided into about 200 PU bands. Implementing a CR system based on the proposed design would be very well possible with today's hardware. Such initial realizations will additionally give important insights for the refinement of the system design for wide-band implementations.

Another question relates to the feasibility of the partial sensing approach. We have shown analytically in Section 5.2 that energy detection is also possible only using some part of the PU band. Furthermore, there are various studies on spectrum sensing based on detection of pilots in the signal [9, 21, 33], which require only the part of the PU band containing the pilot to be sensed. We have furthermore shown in Section 5.3 that frequency diversity is to be expected (at least for the exemplarily investigated PU systems), which supports the usage of a partial sensing approach. Note that the main result of our investigation in Section 5.3.2 is that in most scenarios the achievable time diversity is expected to be a lot higher than the achievable frequency diversity. This result strongly supports the usage of the partial sensing approach, where always the maximum time diversity is exploited. Whereas this discussion indicates that partial sensing seems to be a reasonable approach from the theoretical point of view, also the hardware constraints need to be considered. In real systems, possibly guard bands are required between the sub-channels used for sensing and the sub-channels used for data transmission. The concept of guard bands is frequently applied in wireless communications and does not create a challenge from the implementation point of view. In GSM guard bands between different providers of 100 kHz are used [131], in CDMA2000 the guard bands are 270 kHz [139] and in UMTS 200 kHz [93]. Based on these examples, guard bands should only have a small, quantitative impact on the partial sensing approach and should not qualitatively change the results presented in this thesis.

Non-contiguous SULs are not yet supported by any wireless systems. One of the reasons for choosing OFDM as an underlying technology for the proposed system design is the current research in the area of dynamic resource allocation in Orthogonal Frequency Division Multiple Access (OFDMA) systems [15, 16]. For next generation LTE (LTE-Advanced [1]) such approaches will be used to dynamically allocate sub-carriers to users based on the sub-carriers quality, resulting in communication links with non-contiguous sub-carriers. Thus, we believe that OFDM systems supporting non-contiguous spectrum allocation will be available in the near future.

The two proposed approaches for secondary QoS maintenance have a fundamental limitation. If the usage of the PU bands covered by the CR system is strongly correlated, the secondary QoS improvement achievable is very limited. This is why our proposed CR system should ideally operate in a wide spectrum range covering diverse PUs. For the 900 MHz example considered above, the redundancy and resource reservation approaches can only be applied if there is at least some variability in the usage of the different PU bands. In [161], it is shown that there seems to be considerable variation at least in the PU bands in between heavily used and idle bands. Furthermore, we have shown in [170] that call interarrival times are exponentially distributed and that the autocorrelation of hourly sequences of call arrivals is very low. Although not conclusive these results suggest independence of call arrivals, which supports the usefulness of the proposed secondary QoS mechanisms for cellular PU bands.

CHAPTER 7

Secondary Quality of Service Support

This chapter proposes performance metrics for secondary QoS support. The CR system design proposed in the previous chapter is evaluated with respect to the supported secondary QoS. Specifically, the impact of variability of available spectrum — quantified by Primary User agility and false positives — on the secondary QoS is investigated. The presented investigation is based on results presented in [42]. Furthermore, a short evaluation is presented, investigating to which degree the obtained results are applicable to other CR system designs. Finally, related work is presented and the chapter is concluded.

7.1 Performance Metrics

We define two performance metrics for the investigations presented in this thesis. The *spectral capacity* is defined as a metric quantifying the degree of offered QoS support. Furthermore, we define the *spectral overhead* as a metric to quantify the overhead required to achieve a certain QoS.

7.1.1 Spectral Capacity

The QoS metric commonly used within the spectrum domain is based on the offered spectral capacity. The focus of the presented work is not on a per-user throughput or the performance of a specific type of application, but rather on the amount and continuity of spectral capacity offered by a Secondary User Link (SUL).

The link capacity (C) is defined as the amount of bandwidth offered by a SUL during a certain time period. Assuming a slotted system with a slot length of t_{slot} , the maximum link capacity (C_{max}) per slot is defined as

$$C_{\text{max}} = b_{\text{sul}} \cdot t_{\text{slot}} , \quad (7.1)$$

where b_{sul} is the SUL bandwidth. Remember that in contrast to traditional links, the spectral capacity offered by SULs may vary. Some sub-channels might get occupied by the PU and, thus, have to be excluded from the SUL or the usage has to be interrupted to perform sensing. In both cases, the offered capacity is temporarily reduced.

Let C_i denote the link capacity in slot i . Then the normalized average link capacity over N slots can be computed as

$$C_{\text{avg}} = \frac{1}{N \cdot C_{\text{max}}} \sum_{i=1}^N C_i . \quad (7.2)$$

In addition to the average link capacity we investigate α quantiles of the link capacity (C_α). An α -quantile is essentially a specific point taken from the CDF. C_α specifies the minimum link capacity which is achieved in $\alpha \cdot 100$ percent of the time.

7.1.2 Spectral Overhead

Maintaining a certain secondary QoS is usually not for free. We define *spectral overhead* as a metric quantifying the cost of secondary QoS support.

Additional spectral resources used to maintain the link capacity are one part of the spectral overhead. For the proposed system design, those additional resources are the redundant / reserved sub-channels for a SUL, as introduced in Section 6.4. Increasing the number of redundant / reserved sub-channels of a SUL (X), increases the secondary QoS, but also increases the spectral overhead.

The other part of the spectral overhead for secondary QoS support is defined by the spectral resources needed for the control information exchange for QoS support and link maintenance. However, since the detailed protocol designs are beyond the scope of this thesis, the protocol spectral overhead is not considered.

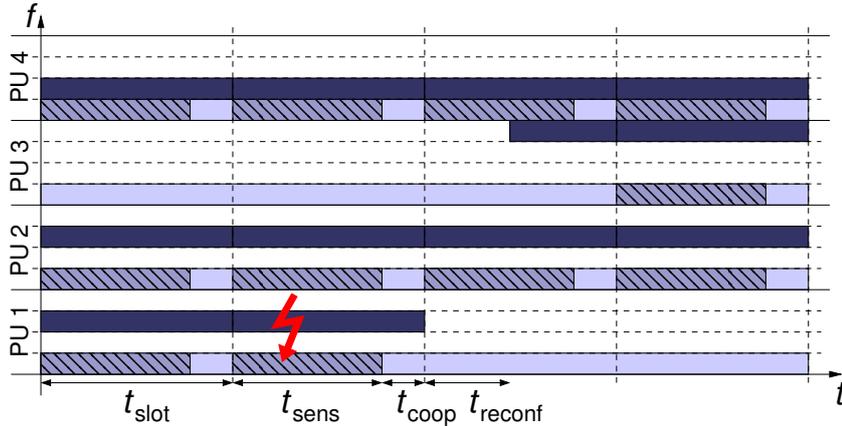


Figure 7.1: Secondary QoS support system model. The dark blue sub-channels are used for the SUL, the light blue are reserved for sensing. The shaded areas represent the local sensing periods.

7.2 Impact of Primary User Agility and False Positives on Secondary QoS

In this section, we evaluate the impact of link reconfigurations on the QoS of a CR system as described in Chapter 6. We quantify the QoS of a SUL in terms of available link capacity (C). During the reconfiguration process, the parts of the SUL to be reconfigured cannot be used for data transmission, and, thus, temporarily reduce the link capacity. The number of link reconfigurations mainly depends on the variability of available spectrum, i.e., the PU agility and the probability of false positives.

7.2.1 System Model

The assumed system model is based on the system design described in Chapter 6. For this investigation, however, we consider a system without specific secondary QoS support mechanisms, i.e., without the redundancy or resource reservation approach. If some sub-channels have to be removed from the SUL, the link capacity is reduced accordingly until new sub-channels are acquired for the SUL, which requires a time period of t_{reconf} .

A corresponding example is shown in Figure 7.1. Here, we show four PU bands (labeled PU 1 to PU 4) and a SUL with $N_{\text{sub}} = 3$ sub-channels (dark blue sub-channels) scattered over the PU bands. In the second slot, a PU is detected in PU band 1 so that the corresponding sub-channel has to be excluded from the SUL. In the following slot, the link capacity is reduced by one sub-channel for the duration of t_{reconf} , after which a new sub-channel (located in PU band 3) is added to the SUL. For the presented performance analysis, we assume $t_{\text{reconf}} \leq t_{\text{slot}}$ and that t_{reconf} does not depend on the number of sub-channels to reconfigure.

Primary User Model

We exemplarily investigate three different PU types, representing different activity patterns which cover a wide range of today's typical wireless communication: TV band PUs with a very “sluggish” activity pattern (rare access times and long usage durations), cellular telephony having a more “agile” activity pattern (frequent access times and short usage duration), and cellular data traffic representing a very agile activity pattern.

The PU activity pattern is described by an arrival rate (λ_{pu}) and a usage duration (μ_{pu}), which characterize a single user within a PU band. Depending on the type of PU, a different number of users can be present within one PU band. Within a TV band, there is only one user per band (one TV channel uses one band exclusively). In the case of cellular telephony and data traffic we assume — based on GSM [43] — up to 8 users (calls) per PU band.

For the latter two PU types resource allocation is assumed to be frequency efficient. Newly arriving users are assigned to “not full” PU bands until the maximum number of users per band is reached. If a user terminates, the PU system may relocate other users internally to use as few PU bands as possible.

Based on Little's Law [100], the average PU system load (L_{pu}), i.e., the average fraction of PU bands occupied by PUs can be calculated to

$$L_{\text{pu}} = \lambda_{\text{pu}} \cdot \mu_{\text{pu}} . \quad (7.3)$$

Modeling False Positives

The sensing process determines in each slot which PU bands are occupied and which are available. The details of the sensing process are not considered for this investigation. At the end of each slot, the sensing process provides for each PU band, whether a PU was detected or not.

This can be regarded as an *estimation* of the PU system load (L_{pu}). However, due to false positives and false negatives in the sensing process, the estimate might not be accurate and differ from L_{pu} . The sensing process has no way to differentiate whether a PU band is really occupied by its owner, or whether the detection is a false positive. Thus, to the sensing process, false positives are a simple increase in PU system load.

In the following work, the probability of false positives (P_{fp}) is modeled as an increase of PU system load. Let L_{sens} be the PU load that was sensed by the sensing process and L_{pu} the real PU load, then P_{fp} can be calculated as

$$P_{\text{fp}} = \frac{L_{\text{sens}} - L_{\text{pu}}}{1 - L_{\text{pu}}} , \quad (7.4)$$

under the assumption that false negatives are negligible.

7.2.2 Performance Analysis Setup

The presented work investigates the influence of the PU behavior as well as of selected CR system parameters on the secondary QoS. Furthermore, the influence of false positives of the sensing process on the secondary QoS is investigated.

Performance Metrics

The performance metric used is the link capacity as introduced in Section 7.1. Specifically, the average link capacity (C_{avg}) and the $\alpha = 0.9999$ and $\alpha = 0.95$ α quantiles of the link capacity (C_α) are investigated. The spectral overhead is not considered in this investigation. For the influence of the spectral overhead refer to Chapter 8.

For the investigations of the influence of false positives on the link capacity (Section 7.2.4), the link capacity achieved in the case of no false positives ($P_{\text{fp}} = 0$) is used for normalization instead of C_{max} . This enables the quantification of only the influence of false positives on the link capacity without considering the impacts of the reference load (reconfigurations caused by true positives) and the influence of (potentially sub-optimally chosen) system parameters.

Method of Evaluation

The simulation is mainly programmed in pure C++ using the *Scythe* library [123]. For the evaluation, Java and R [56] are used.

First, for each of the three PU activity patterns, trace files containing the primary usage for different PU loads are generated. For each activity pattern / PU load combination several runs using different seeds are produced. The total time of one simulation run is set to $T = 100\,000$ s for the telephony and the data traffic PU and to $T = 1\,000\,000$ s for the TV band PU. The generated trace files are then used to simulate the achievable link capacity using different SUL parameter combinations.

The number of runs averaged for the investigation of one SUL parameter combination is chosen such that the average link capacity results are within a 0.25 % interval with a confidence of 95 %.

Parameterization

The simulated system covers B MHz partitioned into $N_{\text{pu}} = 100$ PU bands. The bandwidth covered by one PU band (and thus the total bandwidth B) depends on the PU type (6–7 MHz for TV bands, 200 kHz for GSM). However, note that B has no influence on the presented results.

The telephony PU activity pattern is based on the model developed in [170] and presented in Appendix A with a usage duration (μ_{pu}) following a lognormal distribution and a mean of $E[\mu_{\text{pu}}] = 138$ s. For comparability reasons, the same model is used for the data traffic PU with a reduced mean of $E[\mu_{\text{pu}}] = 2.5$ s. For the TV band PU a normal distributed usage duration with a mean of $E[\mu_{\text{pu}}] = 10\,000$ s is assumed. For all PUs, an exponentially distributed arrival rate process is assumed whose intensity (λ_{pu}) depends on the PU system load.

The underlying OFDM platform of the CR system has $N_{\text{ofdm}} = 2048$ sub-carriers (motivated by IEEE 802.16 [80]) equally spread across the 100 PU bands. There is one single SUG using a SUL with a bandwidth of $b_{\text{sul}} = 20 \cdot b_{\text{sc}}$, i.e., consisting of 20 sub-carriers. The number of sub-channels of a SUL (N_{sub}) is varied between $N_{\text{sub}} =$

Table 7.1: Important parameters and values used for the secondary QoS evaluation

Acronym	Description	Values used
t_{slot}	slot length	$\{0.125, 0.25, 0.5, 1, 2, 4, 8\}$ s
t_{reconf}	link reconfiguration time	$\{0.125, 1\}$ s
N_{sub}	number of sub-channels of a SUL	$\{1, 2, 4, 5, 10, 20\}$
N_{sc}	number of OFDM sub-carriers per sub-channel	$\{20, 10, 5, 4, 2, 1\}$
b_{sul}	SUL bandwidth	$N_{\text{sub}} \cdot N_{\text{sc}} = 20$
L_{pu}	PU system load	$\{5, 50, 65\}$ %
P_{fp}	probability of false positives	0 – 100 %

$\{1, 2, 4, 5, 10, 20\}$ resulting in $N_{\text{sc}} = \{20, 10, 5, 4, 2, 1\}$ sub-carriers per sub-channel, respectively.

Note that N_{sc} is only a scaling factor assuring the comparability of the results and is only limited by the number of sub-carriers per PU band and the investigated N_{sub} range. A SUL with, e.g., 40 sub-carriers would give the same results as presented here by doubling N_{sc} (except that, for the parameters chosen in this investigation, it could not be realized for $N_{\text{sub}} = 1$).

So far there are no specifications on the maximum interference time (t_{max}), except for the TV bands in the U.S., where t_{max} is set to two seconds. We, thus, set the upper bound of t_{max} such that the probability that a PU appears and disappears within the same slot is negligible. In this work results for a slot length of $t_{\text{slot}} = \{0.125, 0.25, 0.5, 1, 2, 4, 8\}$ s and a link reconfiguration time of $t_{\text{reconf}} = \{0.125, 1\}$ s (limited to the cases of $t_{\text{reconf}} \leq t_{\text{slot}}$) are presented. An overview on the most important parameters of our investigation and their commonly used values (unless stated otherwise) is given in Table 7.1.

7.2.3 Link Capacity Without False Positives

In this section results quantifying the link capacity without considering false positives are presented. This means that a perfect sensing process is assumed. The results give insights into the influence of different SUL parameters, the PU activity pattern, and the PU system load on the link capacity.

Average Link Capacity

Comparing the three solid graphs in Figure 7.2 shows the influence of different PU activity patterns on the average link capacity (C_{avg}): the more agile the PU, the more reconfigurations are necessary, resulting in a decreased C_{avg} . The telephony PU and the TV band PU graphs show that up to a load of $L_{\text{pu}} = 65\%$ there is no visible degradation of C_{avg} . For both PUs, reconfigurations are extremely rare so that they do not significantly influence C_{avg} . For the data traffic PU, C_{avg} degrades as L_{pu} increases. A

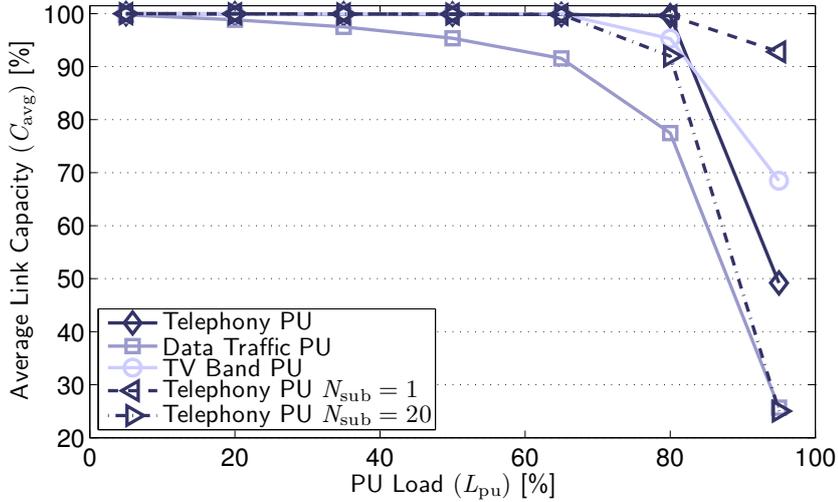


Figure 7.2: Impact of the PU system load (L_{pu}) and the PU activity pattern on the average link capacity (C_{avg}) for $N_{sub} = 10$ and $t_{slot} = t_{reconf} = 0.125$ s.

higher load causes an increased number of reconfigurations, which results in a decreased average link capacity.

For all three PUs the average link capacity is significantly decreased for very high PU loads. This performance degradation is due to the inability to allocate all requested sub-channels of the SUL, so-called blockings. If a SUL, e.g., requires 10 sub-channels, but the sensing process only reports that 8 PU bands are not occupied by PUs at the moment, the SUL can only use 8 of its required 10 sub-channels resulting in a decreased C_{avg} . Note that, for a load of $L_{pu} = 80\%$, the telephony PU almost shows no blockings whereas the TV band PU already has a significant C_{avg} drop due to blockings. This is the result of the different PU structures. The 8 users per PU band for the telephony PU result in an overall reduced variance of the load, which results in fewer blockings. Comparing the 3 graphs for the telephony PU in Figure 7.2 shows the influence of N_{sub} on the blockings. A higher number of sub-channels requires more PU bands to be available (since only one sub-channel is allowed per PU band). Thus, for high loads, a smaller N_{sub} shows better performance. In order to be able to better analyze the influence of reconfigurations, we further concentrate on cases where L_{pu} does not result in blockings for our analysis.

Figure 7.3 shows that N_{sub} has no impact on C_{avg} in case there are no blockings. Statistically, a SUL consisting of 20 small sub-channels each having a width of 1 sub-carrier needs to be reconfigured 20 times more often than a SUL consisting of one large sub-channel of width 20 sub-carriers resulting — on average — in the same reconfiguration-based capacity loss. Figure 7.3 also shows that in case of the telephony PU, there is only a minor difference between the average link capacity for $L_{pu} = 65\%$ and $L_{pu} = 5\%$, as compared to the data traffic PU. This can be attributed to the much higher number of

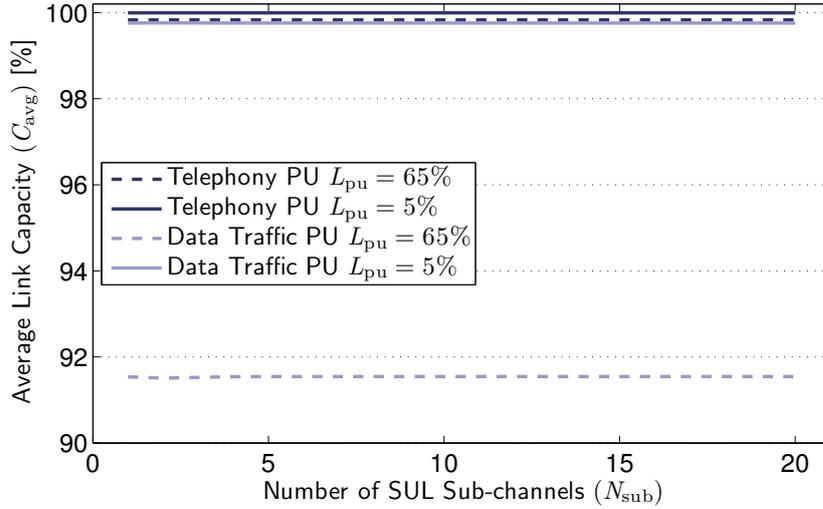


Figure 7.3: Impact of the number of sub-channels of a SUL (N_{sub}) on the average link capacity (C_{avg}) for $L_{\text{pu}} = 65\%$ and $t_{\text{slot}} = t_{\text{reconf}} = 0.125$ s.

active / passive transitions of the data traffic PU. Note that the difference for the TV band PU is even smaller and, thus, not shown.

Figure 7.4 shows that t_{slot} has no influence on the average link capacity and — as would be expected — that longer reconfiguration periods (t_{reconf}) result in a higher capacity loss and, thus, in a lower C_{avg} . Due to the very rare reconfigurations, however, the impact of t_{reconf} is very small (note the large scale of the y-axis) and hardly visible in case of the TV band PU.

Summarizing, given a reasonable PU load such that the bandwidth requirements of the SUL can be satisfied, C_{avg} mainly depends on the PU activity pattern and on the time required for reconfiguration (t_{reconf}). The number of sub-channels of a SUL (N_{sub}) and t_{slot} have no influence on C_{avg} . The PU system load (L_{pu}) only has a significant influence on C_{avg} in case of the data traffic PU.

Quantiles

In contrast to the average link capacity, the number of sub-channels of a SUL (N_{sub}) and the slot length (t_{slot}) have an impact on the α quantile of the link capacity (C_α). The impact depends on α and the PU activity pattern (see Figure 7.5) and is mainly influenced by how many sub-channels need to be reconfigured *simultaneously*.

Looking at $t_{\text{slot}} = 0.125$ s in Figure 7.5(c), one can see that C_α increases from 0 % (1 sub-channel) to about 90 % (20 sub-channels). Spreading the Secondary User Link (SUL) over more sub-channels (and, thus, more PU bands) decreases the probability that multiple sub-channels have to be reconfigured *simultaneously*, resulting in a higher C_α . Fixing $N_{\text{sub}} = 20$ sub-channels in Figure 7.5(c) shows that C_α is continuously decreasing

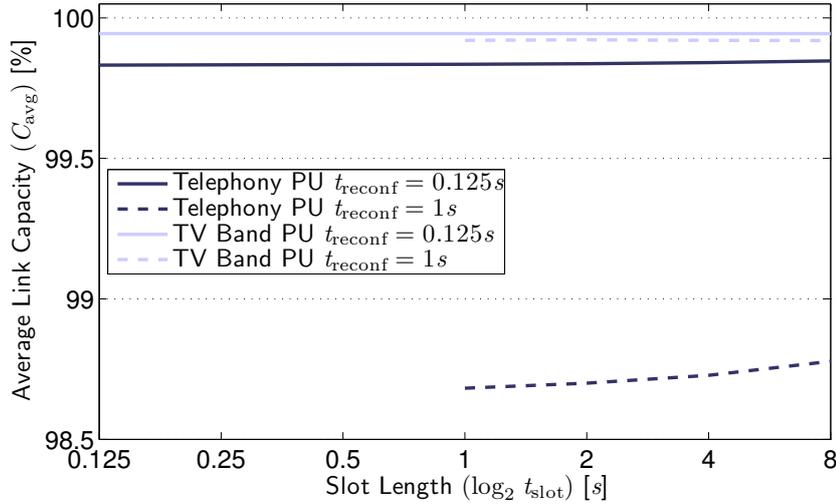


Figure 7.4: Impact of the slot length (t_{slot}) and link reconfiguration time (t_{reconf}) on the average link capacity (C_{avg}) for $N_{\text{sub}} = 10$ and $L_{\text{pu}} = 65\%$.

as t_{slot} increases. The longer t_{slot} , the higher the probability that multiple sub-channels have to be reconfigured *simultaneously* which decreases C_{α} .

The effects described above very much depend on the considered PU. The more agile the PU, the more pronounced are the effects. This can be seen comparing Figure 7.5(a) and Figure 7.5(c). For the very sluggish TV band PU (Figure 7.5(e)) the described effects are only visible comparing $N_{\text{sub}} = 10$ and $N_{\text{sub}} = 20$ sub-channels. For $N_{\text{sub}} < 10$, C_{α} even reaches 100% and outperforms larger N_{sub} values. The reason for this is that even though the fraction of the SUL which has to be reconfigured simultaneously increases, the *total* time required for reconfiguration is still below $(1 - \alpha) \cdot 100\%$ for the very sluggish TV band PU, thus achieving superior C_{α} values.

Which value α to choose depends on the requirements and the sensitivity of the applications using the SUL. In Figures 7.5(b), 7.5(d), and 7.5(f) additional results for $\alpha = 0.95$ are shown. Whereas for the data traffic and TV band PU (Figure 7.5(b) and Figure 7.5(f)) the results only differ quantitatively, for the Telephony PU (Figure 7.5(d)) the same effect as in Figure 7.5(e) can be seen. Thus, for the telephony PU the optimal parameters heavily depend on the chosen α quantile. Considering $\alpha = 0.9999$, a large N_{sub} and a small t_{slot} achieve the maximum C_{α} ; the reverse is true for $\alpha = 0.95$.

7.2.4 Influence of False Positives on Link Capacity

Having understood the influence of the PU activity pattern and selected parameters such as N_{sub} and t_{slot} on the link capacity, we now investigate the influence of false positives on the link capacity. Remember that in order to achieve a reliable sensing process, the price (especially for early, low cost CR deployments) is often a high probability of false positives. We are, thus, interested in whether a certain secondary QoS quantified by the

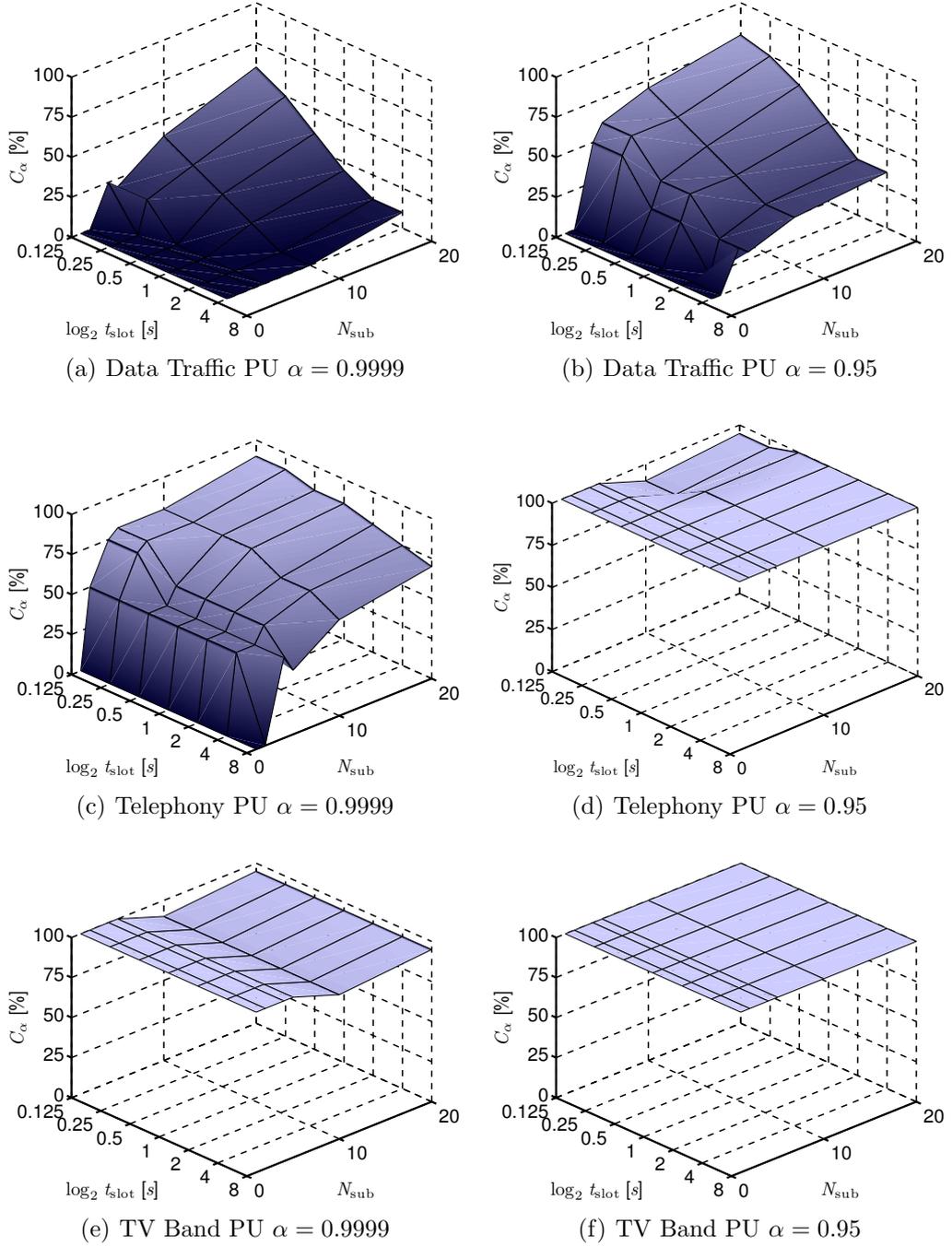


Figure 7.5: The α quantile of the link capacity (C_α) for $\alpha = 0.9999$ (left column) and $\alpha = 0.95$ (right column). The plots are shown for $t_{\text{reconf}} = 0.125$ s and $L_{\text{pu}} = 65\%$ except for Figures (e) and (f) where $L_{\text{pu}} = 50\%$ in order to avoid the influence of blockings.

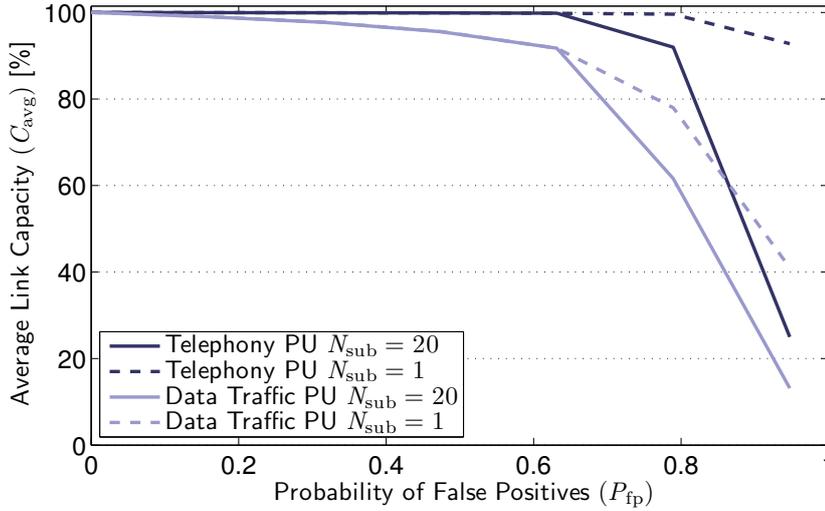


Figure 7.6: Impact of the probability of false positives (P_{fp}) on the average link capacity (C_{avg}) for $L_{pu} = 5\%$ and $t_{slot} = t_{reconf} = 0.125$ s.

average link capacity (C_{avg}) and the α quantile of the link capacity (C_α) can also be maintained in case of a high probability of false positives (P_{fp}).

Average Link Capacity

Figure 7.6 shows the influence of false positives on the average link capacity (C_{avg}), assuming an actual PU system load of $L_{pu} = 5\%$. For the telephony PU, C_{avg} only starts to decrease for $P_{fp} > 63\%$, which is basically caused by the very high load ($L_{sens} > 65\%$) *perceived* by the SUG, resulting in a high number of blockings, as described in Section 7.2.3. As already elaborated, for such high loads fewer sub-channels achieve a higher C_{avg} . The average link capacity in case of the data traffic PU shows a similar behavior. Since reconfigurations have a higher impact in case of very agile PUs, there is a slight decrease of the average link capacity even for low false positives rates. The effects for the TV band PU are similar to the telephony PU and are, thus, not shown.

In Figure 7.7, the influence of different actual PU system loads is compared. The bottom two graphs ($L_{pu} = 50\%$) show that for a higher actual load fewer false positives can be tolerated, as compared to a lower load (upper two graphs, $L_{pu} = 5\%$). Furthermore, Figure 7.7 shows (as already seen in Section 7.2.3) that a longer t_{reconf} results in a degraded C_{avg} ; however, the difference between $t_{reconf} = 0.125$ s and $t_{reconf} = 1$ s is marginal.

Quantiles

The influence of P_{fp} and N_{sub} on the α quantile of the link capacity can be seen in Figure 7.8. Generally speaking, a larger N_{sub} achieves better results, except for the

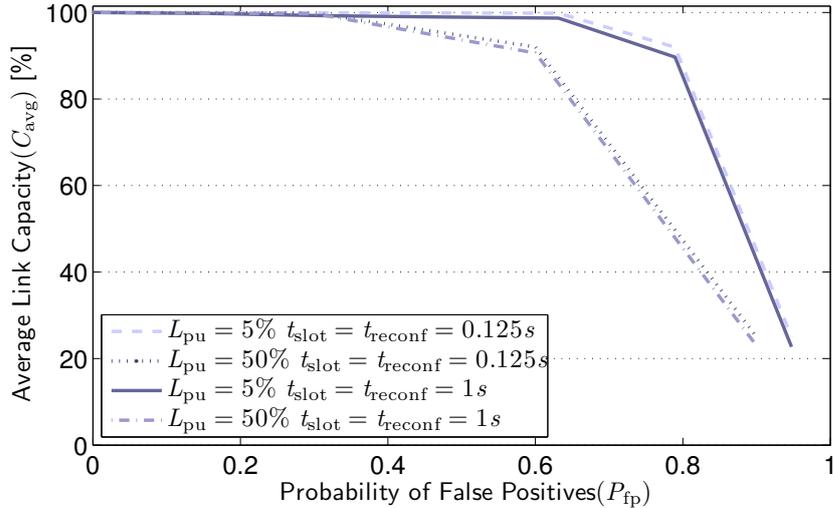


Figure 7.7: Impact of the actual PU system load (L_{pu}) and the link reconfiguration time (t_{reconf}) on average link capacity (C_{avg}) for $N_{sub} = 10$.

TV band PU (Figure 7.8(c)) where also one or two sub-channels achieve very good results. This is in accordance to our previous observations of Section 7.2.3. Similar to the results for the average link capacity (Figure 7.6) a very high C_α can be maintained for $P_{fp} < 50\%$.

7.3 QoS Support in Dynamic Frequency Hopping Systems

In this section we give some intuition on how link reconfigurations impact the secondary QoS support in other CR system designs. Specifically, we focus on Dynamic Frequency Hopping (DFH) CR systems. We chose DFH systems mainly for two reasons. First, DFH systems offer a periodic sensing process, which does not require to interrupt usage of a SUL for spectrum sensing, thus already providing some basic secondary QoS. Second, DFH has been proposed and is considered as an alternative operation modus within the IEEE 802.22 standardization and, thus, has a high practical relevance.

As in the previous section, the performance metric investigated is the link capacity, more precisely the average link capacity (C_{avg}) and the α quantile of the link capacity (C_α). Note however, that we do not provide a complete performance analysis, but rather point out some intuitive results based on the observations made in the previous section.

7.3.1 System Model

The system model considered is based on the CR system design presented in Chapter 6. Instead of a partial sensing approach, DFH is assumed to perform a periodic sensing

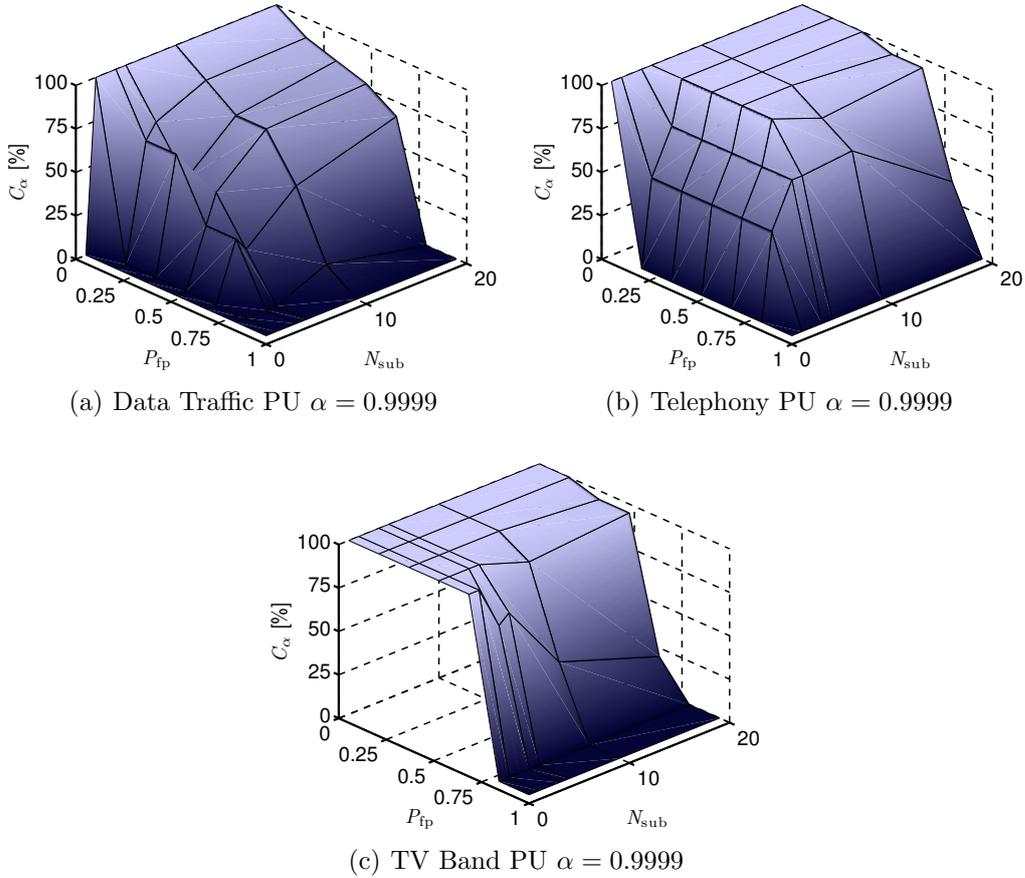


Figure 7.8: Influence of the probability of false positives (P_{fp}) on the α quantile of the link capacity (C_α) for $t_{slot} = t_{reconf} = 0.125$ s and $L_{pu} = 5\%$.

process. Furthermore, since DFH was designed for usage of contiguous SULs, only contiguous SULs are considered. An example showing the resulting system model is shown in Figure 7.9. It shows four PU band (PU 1 to PU 4) and a SUL consisting of $N_{sub} = 1$ sub-channel with $N_{sc} = 3$ sub-carriers located in PU band 1. Before hopping to a new PU band, periodic sensing is performed on that PU band taking a total time of $t_{sens} + t_{coop}$. In the example, a PU appears in PU band 1 in the fourth slot, such that the PU band cannot be used anymore by the SUL in the following slot. Consequently, the SUL is reconfigured and switched to PU band 3, which obviously has to be previously validated to ensure its availability. This procedure takes a time period of t_{reconf} , during which the link capacity drops to zero.

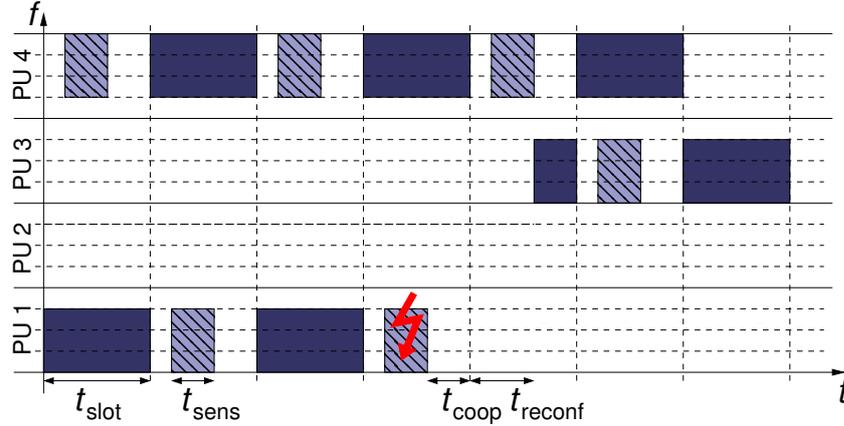


Figure 7.9: Secondary QoS support system model for Dynamic Frequency Hopping (DFH). The dark blue sub-channels are used for the SUL, and the shaded areas represent the local sensing periods.

7.3.2 Performance Assessment

Comparing the DFH system model with the one used in the previous section, we observe that the DFH system model is essentially a special case of the system model used for the investigation in Section 7.2. Using $N_{\text{sub}} = 1$ and $N_{\text{sc}} = 20$ in the previous section results in a SUL with only one big sub-channel, which is exactly what is assumed for the DFH system model. We can, thus, assess the secondary QoS support in case of the DFH approach by looking at the results for $N_{\text{sub}} = 1$ from the previous section. C_{avg} is not influenced significantly by reconfigurations. Only in case of the very agile data traffic PU, C_{avg} is slightly decreased for a high L_{pu} (see Figure 7.2). As the partial sensing approach, the DFH approach can tolerate a very high probability of false positives (P_{fp}) without a significant influence on the average link capacity (see Figure 7.6). Looking at C_{α} for $N_{\text{sub}} = 1$ (Figures 7.5 and 7.8), in contrast, shows that the DFH approach performs very poorly and is significantly influenced by reconfigurations and, thus, also by false positives. The reason is the same as already indicated above. If a PU is detected, the *entire* SUL always has to be reconfigured and the link capacity drops to zero. In other words, the diversity gain through spreading the SUL over multiple PU bands as with non-contiguous SULs is lost. Thus, in order to not only maintain a good average link capacity but also a good α quantile of the link capacity, additional mechanisms as e.g., proposed in Section 6.4 are required for the DFH approach.

7.4 Related Work

To the best of our knowledge, the impact of link reconfigurations on the SU performance has so far not been investigated. Reconfigurations of a CR system are investigated in [154]; however, the focus is put on reconfiguring physical layer parameters (e.g., bit

rate, transmit power, FEC) while considering an interfering noise source within one single PU band. Thus, the influence of spectral agility of the SU as well as the advantage of spreading the secondary communication across multiple PU bands are not taken into account. Furthermore, a simple adjustment of physical layer parameters, in case a PU appears, can potentially create harmful interference. Other studies, such as [30], examine SU systems benefiting from spectral agility; however, they do not consider the impact of link reconfigurations on the SU performance. Another advantage of our work is that we investigate a wide range of different PU activity patterns.

7.5 Conclusions

In this chapter we have investigated the **QoS support for secondary communication** for a CR system based on the design presented in Chapter 6. Using the link capacity as a secondary QoS metric, we have quantified the influence of link reconfigurations on the link capacity of an OFDM-based Secondary User Link (SUL). Specifically, we have investigated the influence of the link reconfiguration time (t_{reconf}), the slot length (t_{slot}) and the number of sub-channels of a SUL (N_{sub}). We have shown results for scenarios with a very diverse variability of available spectrum. The variability of available spectrum is the main influencing factor for link reconfigurations and is mainly quantified by the Primary User (PU) activity patterns and the probability of false positives (P_{fp}) of the sensing process. We, thus, investigated a wide range of Primary User (PU) activity patterns with varying false positive rates.

We show that — within the set of chosen parameters — link reconfigurations only have a small impact on the average link capacity (C_{avg}), which does not depend significantly on the chosen SUL parameters. Only for the very agile data traffic PU a significant drop of C_{avg} could be observed. The optimal parameter setting for the α quantile of the link capacity (C_{α}) depends on the PU, and on the selected α quantile. Whereas in general a large N_{sub} and a small t_{slot} achieves superior results, the opposite might be true for a smaller α quantile or a less agile PU.

Our results also show that a probability of false positives (P_{fp}) of up to 50% has no significant influence on C_{avg} as well as C_{α} for all investigated PU activity patterns. Depending on the PU agility even higher false positive rates can be tolerated.

There are two main conclusions that we can draw for our proposed system design from the results presented in this chapter:

1. The proposed CR system design is able to operate in PU bands with a wide range of PU activity patterns. Specifically, a certain QoS support for the SUL can even be supported in case of very agile PUs.
2. Even for a really high probability of false positives our proposed CR system design can maintain a certain secondary QoS. We can, thus, use low complexity energy detection based sensing processes as described in Chapter 5 without significantly reducing the secondary QoS.

CHAPTER 8

Spectral Efficiency in Cognitive Radio Networks

In the previous chapters of this thesis it was shown that reliable Primary User (PU) detection (Chapter 5) and secondary QoS maintenance (Chapter 7) is possible also in PU bands of highly agile PUs. However, for both a certain spectral overhead is required. In this chapter, extending work presented in [164–166], performance metrics to quantify the individual spectral overheads are introduced, followed by a performance analysis of the trade-off between the spectral overhead required for reliable PU detection and for secondary QoS support. Investigations for both a partial sensing and a Dynamic Frequency Hopping-based CR system are shown. Related work and conclusions are presented in the last two sections.

8.1 Performance Metrics

In order to enable an orderly performance analysis of the spectral overhead required by spectrum sensing and secondary QoS maintenance and to be able to judge the spectral efficiency of CR systems, performance metrics quantifying the spectral efficiency are introduced in this section.

For simplicity we consider in the following a spatial area small enough to assure the *unified* availability or non-availability of spectrum for opportunistic re-usage. For such an “elementary area” we define available spectrum in terms of frequency and time availability. To quantify the available spectrum, we conceptually discretize time into slots of length t_{slot} and define available spectrum at a basis of t_{slot} . We differentiate between three notions of available spectrum: theoretically available spectrum (S_{max}), sensed available spectrum (S_{sens}), and effectively available spectrum (S_{eff}).

8.1.1 Theoretically Available Spectrum

As the theoretically available spectrum (S_{max}) we denote the sum of spectrum which is not used by the PU (within a time slot and the elementary spatial unit under consideration). The theoretically available spectrum is the “ground truth” or the “Gods view” and reflects the real availability of spectrum (based on the spectrum usage of the PU). It can be seen as a benchmark to compare the performance of different DSA approaches. Up to now, there are only very limited publicly available research results analyzing the theoretically available spectrum. Part of the reason is that in-network data is needed, which spectrum owners often are not willing to share. One of the few exceptions publicly available was done in the course of this thesis in a cooperation with a big U.S. network provider [171] and is summarized in Section 4.3. Another example is the assessment of unused TV spectrum based on the evaluation of the TV station database for the U.S. [148].

8.1.2 Sensed Available Spectrum

The sensed available spectrum (S_{sens}) is the spectrum *sensed* to be available for secondary usage and is, thus, an estimate of the theoretically available spectrum as discovered using a given sensing approach. How close this estimate comes to the theoretically available spectrum depends on the quality of the sensing process, i.e., the probability of false positives.

Recall that the reliability of the sensing process is to be guaranteed, i.e., the sensing process has to provide some very limited probability of false negatives. Thus, the sensed available spectrum solely depends on the probability of false positives and $S_{\text{sens}} \leq S_{\text{max}}$. It specifies how much of the idle spectrum can be discovered without declaring spectrum used by the PUs as idle. Note however, that the sensed available spectrum does not consider the spectral overhead required for sensing. The spectral overhead is accounted for in the effectively available spectrum (S_{eff}) defined below.

8.1.3 Effectively Available Spectrum

In Chapter 5 we have elaborated that each sensing approach requires some spectral overhead: keeping some spectrum resources unused in order to make sensing possible and also spectral resources needed for exchanging sensing data in the case of distributed sensing. The definition of the sensed available spectrum ignores this need.

Likewise, as investigated in Chapter 7, also the link maintenance process requires spectral overhead to maintain the secondary QoS: backup spectrum / redundancy to compensate for reconfigurations and additional spectrum for signaling (to negotiate the reconfiguration of the SUL within the SUG).

Therefore, we introduce the additional concept of effectively available spectrum (S_{eff}) being the part of the spectrum which actually can be used for secondary data transmission according to the spectrum sensing results achieved on the basis of the used approach for spectrum sensing and SUL reconfiguration. To be able to separately quantify the sensing spectral overhead, we also introduce the effectively available spectrum after spectrum sensing ($S_{\text{eff}}^{\text{sens}}$) as an intermediate metric.

The effectively available spectrum quantifies the efficiency of a given secondary usage approach in terms of how much of the unused spectrum could be effectively recovered for secondary usage. Additionally, the effectively available spectrum enables us to quantify the tradeoff between the sensing overhead and the overhead required for secondary QoS support.

8.2 Spectrum Sensing and Secondary QoS Support Efficiency Tradeoff

In this section we analyze the spectral efficiency of the CR system design proposed in Chapter 6. It has been shown in the previous chapters of this thesis that for the proposed system design, reliable PU protection and secondary QoS support can be achieved. Here, we investigate the spectral overhead required to achieve reliable PU protection and secondary QoS support. More precisely, we are interested in the *tradeoff* of where to spend the spectral overhead. Putting a lot of effort into the sensing process, i.e., using a lot of sensing spectral overhead, not only assures reliable PU protection but also results in a relatively low probability of false positives (P_{fp}). A low P_{fp} , in turn, results in a low spectral overhead required for secondary QoS support since the probability of link reconfigurations is reduced. Spending only very little spectral overhead for the sensing process, on the other hand, results in a high P_{fp} (and, thus, in a reduced S_{sens}), and additionally requires more spectral overhead for secondary QoS support due to the increased number of link reconfigurations.

8.2.1 System Model

The system model is based on the CR system design presented in Chapter 6. A single SUG is considered, and m CRs of the SUG participate in the distributed sensing process.

Table 8.1: Important parameters for the calculation of the spectrum sensing spectral overhead.

Acronym	Description
M_t	number of sensing samples in the time domain
M_b	number of sensing samples in the frequency domain
$M = M_t \cdot M_b$	total number of sensing samples
N_{sc}	number of OFDM sub-carriers per sub-channel
b_{sc}	OFDM sub-carrier bandwidth
$b_{sub} = N_{sc} \cdot b_{sc}$	sub-channel bandwidth
$b_{coh} = b_{sub}$	coherence bandwidth
$b_{sens} = M_b \cdot b_{coh}$	sensing bandwidth
t_{coh}	coherence time
$t_{sens} = M_t \cdot t_{coh}$	local sensing time

The SUG operates in a spectrum range covering B MHz divided into N_{ofdm} equally sized OFDM sub-carriers. A SUL is constructed out of N_{sub} sub-channels and each sub-channels consists of N_{sc} sub-carriers resulting in a sub-channel bandwidth of $b_{sub} = N_{sc} \cdot b_{sc}$. The maximum number of sub-carriers of one sub-channel is limited by the PU bandwidth ($N_{sc} \leq N_{ofdm}/N_{pu}$). A time-slotted system is assumed, where the periodically repeated time structure is a slot of length t_{slot} , which is set equal to the maximum interference time (t_{max}) and assumed to be a multiple integer of the coherence time ($t_{slot} = t_{max} = N \cdot t_{coh}$). The probability that a PU is using its PU band within t_{slot} is denoted by P_{pu} .

Spectrum Sensing Model

A partial sensing approach is assumed as described in Section 6.2. We use the energy detector presented in [99], which is introduced in Chapter 5 of this thesis. Using this energy detector, the local probability of false positives (P_{fp}) of each sensor is defined as a function of the local sensing time (t_{sens}) and sensing bandwidth (b_{sens}), the diversity in time and frequency, which depends on the coherence time (t_{coh}) and coherence bandwidth (b_{coh}), the target probability of false negatives (P_{fn}), and the received SNR (γ) as shown in Equations (5.13) and (5.14) of Section 5.2.

The number of sensing samples (M) used for the local diversity of the sensing result (defined in Equation (5.9) of Section 5.2) can be split into the number of sensing samples in the time domain (M_t) and number of sensing samples in the frequency domain (M_b), to define the diversity in the time and frequency domain separately, resulting in $M = M_b \cdot M_t$. Using M_t and M_b , the local sensing time can be defined as $t_{sens} = M_t \cdot t_{coh}$ and the sensing bandwidth as $b_{sens} = M_b \cdot b_{coh}$. Without loss of generality we assume that the sub-channel bandwidth is equal to the coherence bandwidth ($b_{sub} = b_{coh}$).

A distributed sensing approach logically OR'ing the individual local sensing results is assumed as described in Section 5.2. The distributed sensing results are exchanged using

a perfect reporting channel and are logically OR'ed at a central controller. We assume the time for sensing results exchange and fusion to be linearly dependent on the number of cooperating sensors ($t_{\text{coop}} = m \cdot \alpha$, with α being a time constant). The probability of false negatives of the SUG (\overline{P}_{fn}) and the probability of false positives of the SUG (\overline{P}_{fp}) are calculated as shown in Equation (5.1) and (5.2) of Section 5.2, respectively.

As spectral overhead only the local spectrum sensing spectral overhead is considered. As defined in Section 6.2.1 the sensing bandwidth is never used for data transmission in the partial sensing approach. The local spectrum sensing spectral overhead can, thus, be calculated as $t_{\text{slot}} \cdot b_{\text{sens}}$. For the distributed sensing process the protocol spectral overhead is not considered and subject to future work. Note however, that the distributed sensing time (t_{coop}) and, thus, the number of cooperating sensors (m) have an influence on the local sensing time, i.e., the available time diversity and probability of false positives (remember that the distributed sensing time depends on the number of cooperating sensors and computes to $t_{\text{coop}} = m \cdot \alpha$, with α being a time constant). Consequently, the distributed sensing time has an indirect influence on the spectrum sensing spectral overhead. The most important parameters for the calculation of the spectrum sensing spectral overhead are summarized in Table 8.1.

Secondary QoS Maintenance Model

Link reconfiguration and secondary QoS maintenance are based on the CR design introduced in Chapter 6. At the end of the distributed sensing process each SUG member knows which sub-channels to exclude from the SUL. To determine which new sub-channels to add to the SUL a time period of t_{reconf} is required.

We investigate both the redundancy and the resource reservation approaches to maintain a certain secondary QoS as described in Section 6.4. In the redundancy approach, sub-channels on which a PU was detected are immediately excluded from the SUL, such that the link capacity is reduced until new sub-channels are added. To compensate for this reduction, redundant sub-channels are added to the SUL. According to Section 6.4.1 and Figure 6.5 the local sensing time computes to

$$t_{\text{sens}} = t_{\text{slot}} - t_{\text{coop}} . \quad (8.1)$$

In the resource reservation approach, the interfered sub-channels are not removed from the SUL until new sub-channels are acquired which puts a much stricter time constraint on t_{reconf} . Furthermore, t_{reconf} also influences the local sensing time as explained in Section 6.4.2 and shown in Figure 6.6. The local sensing time for the resource reservation approach computes to

$$t_{\text{sens}} = t_{\text{slot}} - t_{\text{coop}} - t_{\text{reconf}} . \quad (8.2)$$

As secondary QoS maintenance spectral overhead the number of redundant / reserved sub-channels of a SUL (X) is considered in the presented investigation. The more redundant / reserved sub-channels are used, the bigger is the spectral overhead. The protocol spectral overhead for link reconfiguration and secondary QoS support is

Table 8.2: Important parameters for the calculation of the QoS maintenance spectral overhead.

Acronym	Description
N_{sub}	number of sub-channels of a SUL
X	number of redundant / reserved sub-channels of a SUL
X_{opt}	optimum number of redundant / reserved sub-channels of a SUL
P_{detect}	probability that a PU is detected within its PU band
P_{qos}	probability of violating the QoS requirement of the SUL
P_{pu}	probability of PU presence

not considered and subject to future work. Table 8.2 summarizes the most important parameters required for the calculation of the QoS maintenance spectral overhead.

8.2.2 Performance Analysis

Based on the above system model, the theoretically available spectrum per slot is defined as

$$S_{\text{max}} = (1 - P_{\text{pu}}) \cdot B \cdot t_{\text{slot}} . \quad (8.3)$$

In the following, S_{max} is used as a reference and all spectral efficiency metrics are defined as a fraction of S_{max} .

The sensed available spectrum (S_{sens}) is determined by the probability of false positives of the SUG ($\overline{P_{\text{fp}}}$). It is defined as the theoretically available spectrum (S_{max}) less the reduction in available spectrum due to false positives. Neglecting the influence of $\overline{P_{\text{fn}}}$, S_{sens} can be computed as

$$S_{\text{sens}} = 1 - \overline{P_{\text{fp}}} . \quad (8.4)$$

In the partial sensing process M_{b} sub-channels are always (during the whole slot) reserved for spectrum sensing within each PU band. The spectral overhead of the local sensing process, thus, depends on M_{b} . Since the spectral overhead of the distributed sensing process (the spectrum required for the cooperative sensing protocols) is beyond the scope of this thesis, the effectively available spectrum after spectrum sensing ($S_{\text{eff}}^{\text{sens}}$) can be computed as

$$S_{\text{eff}}^{\text{sens}} = S_{\text{sens}} \left(1 - \frac{M_{\text{b}} \cdot b_{\text{sub}}}{b_{\text{pu}}} \right) . \quad (8.5)$$

In order to calculate the overhead required for the reconfiguration compensation approaches, the number of redundant / reserved sub-channels which are added to a SUL need to be calculated. To assure that the secondary QoS constraint of always maintaining a SUL capacity of N_{sub} sub-channels is satisfied, the number of redundant / reserved sub-channels of a SUL (X) need to be adapted to the number of sub-channels that need to be reconfigured concurrently (within one slot).

The number of sub-channels of a SUL that need to be reconfigured concurrently depends on the probability that a PU was detected. Recall that it does not matter whether

the PU really appeared or whether the sensing process reported a false positive — the respective sub-channel has to be replaced in either case. The probability that a PU is detected within its PU band (P_{detect}) can be computed as

$$P_{\text{detect}} = P_{\text{pu}} \cdot (1 - \overline{P_{\text{fn}}}) + (1 - P_{\text{pu}}) \cdot \overline{P_{\text{fp}}}. \quad (8.6)$$

Since a SUL can maximally contain one sub-channel per PU band, P_{detect} is also the probability that a single sub-channel of a SUL has to be reconfigured. Thus, using P_{detect} , the probability of violating the QoS requirement of the SUL (P_{qos}), i.e., the probability that the capacity offered by the SUL drops below the minimum capacity of N_{sub} sub-channels, can be calculated. Assuming that X additional sub-channels are reserved for the SUL (in case of the reservation approach) or added to the SUL (in case of the redundancy approach), P_{qos} can be calculated using the binomial coefficient as shown in Equation (8.7).

$$P_{\text{qos}} = \sum_{i=1}^{N_{\text{sub}}} \binom{N_{\text{sub}} + X}{X + i} P_{\text{detect}}^{X+i} (1 - P_{\text{detect}})^{N_{\text{sub}}-i} \quad (8.7)$$

The optimum number of redundant / reserved sub-channels of a SUL (X_{opt}) needed to achieve a certain target probability of violating the QoS requirement of the SUL (P_{qos}) can be numerically calculated using Equation (8.7). Using X_{opt} sub-channels for the two maintenance approaches, S_{eff} can be calculated as

$$S_{\text{eff}} = S_{\text{eff}}^{\text{sens}} \left(1 - \frac{X_{\text{opt}}}{N_{\text{sub}} + X_{\text{opt}}} \right). \quad (8.8)$$

8.2.3 Simulation Results

Based on the analysis above we performed Monte-Carlo simulations to investigate the effectively available spectrum (S_{eff}) of our proposed CR system design. We compare S_{eff} with the effectively available spectrum after spectrum sensing ($S_{\text{eff}}^{\text{sens}}$) to evaluate the amount of spectral overhead required for the spectrum sensing and for the secondary QoS maintenance process. We also compare the spectral overhead required for the redundancy and the resource reservation approach.

Parameterization

We use a parameterization which is based on the investigations of the previous chapters wherever applicable. We assume a cellular PU (e.g., UMTS or LTE) with a PU bandwidth of $b_{\text{pu}} = 5$ MHz. Based on the investigations of Section 5.3.1 we vary the coherence bandwidth between $b_{\text{coh}} = 100$ kHz and $b_{\text{coh}} = 500$ kHz and the coherence time between $t_{\text{coh}} = 10$ ms and $t_{\text{coh}} = 100$ ms. Those values correspond to an sub-urban / urban scenario and an average velocity between walking speed and urban vehicular traffic speed. Based on the results of the QoS support investigations in Section 7.2 we choose a SUL with a rather large number of sub-channels ($N_{\text{sub}} = 20$) and a rather short slot length ($t_{\text{slot}} = 0.5$ s). For an overview of the complete parameter setting used for this evaluation refer to Table 8.3. If not stated otherwise, the results are shown for the values in bold.

Table 8.3: Important parameters and values for the efficiency tradeoff evaluation

Acronym	Description	Values used
\overline{P}_{fn}	probability of false negatives of the SUG	0.001
P_{pu}	probability of PU presence	0.2
b_{pu}	PU bandwidth	5 MHz
t_{slot}	slot length	0.5 s
m	number of cooperating sensors	1 – 100
t_{reconf}	link reconfiguration time	0.1 s
t_{coop}	distributed sensing time	$m \cdot 5$ ms
t_{coh}	coherence time	{0.01, 0.1} s
b_{coh}	coherence bandwidth	{100, 500} kHz
N_{sub}	number of sub-channels of a SUL	20
P_{qos}	probability of violating the QoS requirement of the SUL	{0.01, 0.1}
γ	received SNR	–30 dBm

Spectrum Sensing Overhead vs. QoS Maintenance Overhead

In Figure 8.1 we compare the effectively available spectrum after spectrum sensing ($S_{\text{eff}}^{\text{sens}}$) with the overall effectively available spectrum (S_{eff}) for the redundancy approach. From these two figures we can assess and compare the spectral overhead used for spectrum sensing and the spectral overhead used for secondary QoS maintenance. In Figure 8.1(a) we plot $S_{\text{eff}}^{\text{sens}}$ for a varying sensing bandwidth, which quantifies the spectral overhead used for the sensing process. We can clearly see the tradeoff between the sensing bandwidth (b_{sens}) and the effectively available spectrum after spectrum sensing. There exists a certain sensing bandwidth for which the effectively available spectrum after spectrum sensing is maximized. For a very low b_{sens} , the probability of false positives is very high, resulting in a low $S_{\text{eff}}^{\text{sens}}$. With an increasing b_{sens} , the probability of false positives is reduced, resulting in an increased $S_{\text{eff}}^{\text{sens}}$ until the point where the increase of sensing spectral overhead (caused by the increased b_{sens}) is bigger than the decrease of the probability of false positives. Obviously, the optimal sensing bandwidth depends on the system parameters. For a very high time and frequency diversity (light, solid line in Figure 8.1(a)) the optimal b_{sens} is very small, whereas a large sensing bandwidth is optimal for the low diversity case (light, dash-dotted line). Comparing the graphs for $m = 5$ (dark dotted line) and $m = 50$ (dark dashed line) it is interesting to observe, that for less spatial diversity ($m = 5$), $S_{\text{eff}}^{\text{sens}}$ is actually higher than for the high spatial diversity ($m = 50$) case. This results from the decreased local sensing time, since more time is required to exchange and merge the distributed sensing results (remember that in our system model the distributed sensing time linearly increases with m). Thus, too much spatial diversity can result in a decreased $S_{\text{eff}}^{\text{sens}}$.

In Figure 8.1(b) we show results for the same parameters as in Figure 8.1(a) but this time for the overall effectively available spectrum (S_{eff}), i.e., also including the

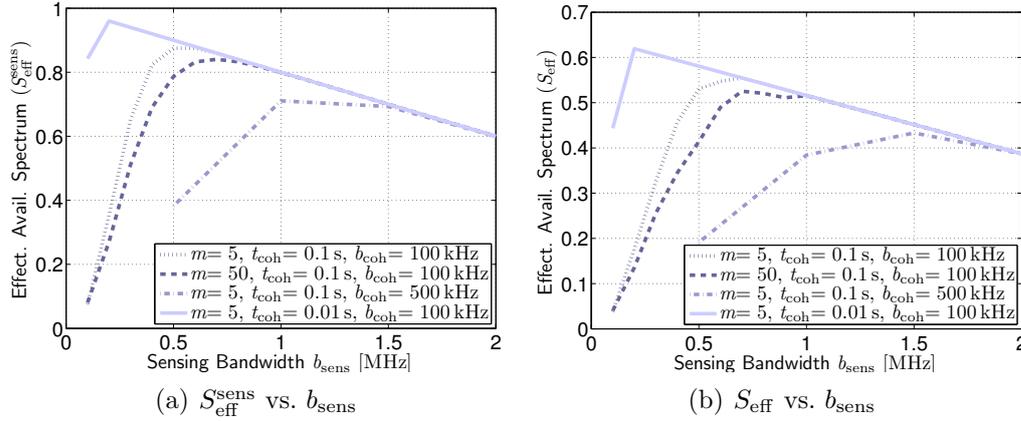
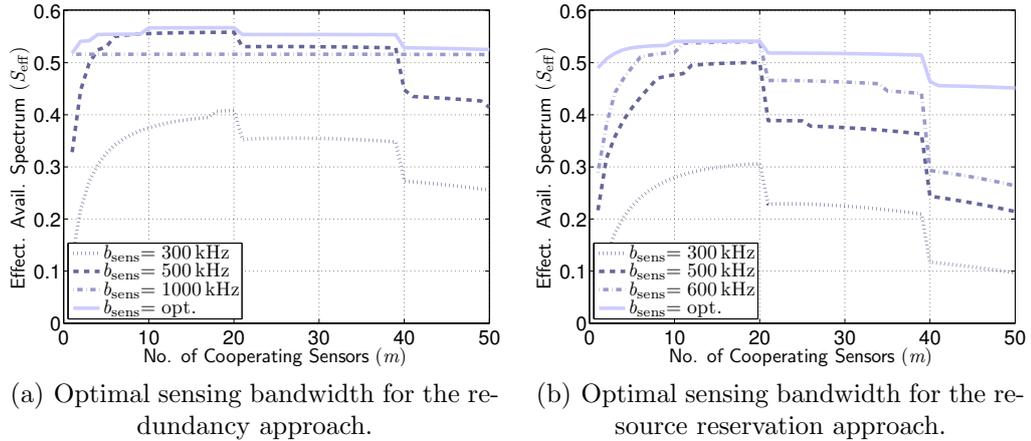


Figure 8.1: Comparison of the spectrum sensing and the QoS maintenance spectral overhead for the redundancy approach.

QoS maintenance spectral overhead. Whereas the two figures are very similar at a first glance, there are several things to be pointed out. First of all, S_{eff} is obviously lower than $S_{\text{eff}}^{\text{sens}}$. Comparing the absolute numbers, the QoS maintenance spectral overhead is always higher than the spectrum sensing overhead. However, we have to keep in mind that the QoS maintenance overhead can be split in two components: the maintenance overhead due to PUs which really appeared (true positives) and the overhead due to the false positives of the sensing process. Whereas the overhead due to false positives can be reduced by a better sensing quality, the overhead due to true positives is fixed. Another thing to point out is that the optimal sensing bandwidth might be slightly shifted in Figure 8.1(b). For the parameters of the dark dotted graph, e.g., the optimal $S_{\text{eff}}^{\text{sens}}$ is achieved for $b_{\text{sens}} = 500 \text{ kHz}$ whereas the optimal S_{eff} is achieved for $b_{\text{sens}} = 700 \text{ kHz}$. This shows the dependency between the spectrum sensing and the QoS maintenance spectral overhead. Using less spectrum for sensing increases the probability of false positives, which results in a higher probability of link reconfiguration and, thus, requires more sub-channels to be reserved in order to ensure the secondary QoS. Conversely, increasing the amount of spectrum used for sensing reduces the effect of false positives but also results in more spectrum blocked for sensing and, thus, not usable for data transmission. Using less spectrum for sensing than the optimum, false positives are the dominating effect on S_{eff} whereas using more than the optimum the spectrum blocked for sensing is the dominating effect on S_{eff} . The optimal point, where S_{eff} is maximized is, in general, not equal to the optimal point for maximizing $S_{\text{eff}}^{\text{sens}}$.

Spectral Efficiency vs. Secondary User Group Size

In the previous figures we have shown that there exists a clear tradeoff between the spectral overhead used for spectrum sensing and the spectral overhead used for secondary QoS maintenance. Furthermore, we have shown that the optimal overhead tradeoff


 Figure 8.2: Investigation of the optimal sensing bandwidth (b_{sens}).

depends on the parameters of the system. We now investigate the scalability of this tradeoff and the spectral efficiency as the size of the SUG (and with this the CRs which can participate in the distributed sensing process) increases. Remember that one design goal for our CR system is scalability. It should support the deployment of initially small CR networks with potentially low spectral efficiency which increases as the network size increases.

In Figure 8.2(a) we show S_{eff} over the number of cooperating sensors (m) for the redundancy approach. Note that m is dependent on the size of the SUG: the bigger the SUG, the more CRs can (but do not have to) participate in the distributed sensing process. We show graphs for different b_{sens} , where the light, solid line shows results using the optimal sensing bandwidth. The general trend we can see for all graphs is an increase of the effectively available spectrum for an increasing number of cooperating sensors up to a certain m , after which the effectively available spectrum decreases again. The S_{eff} decrease for high m was already explained above. At a certain point the decrease of time diversity (due to an increased distributed sensing time) is bigger than the spatial diversity gain. In fact, sharp drops of S_{eff} at regular intervals can be observed. The sharp drops correspond to values of m for which the diversity in the time domain (M_t) is decreased since t_{sens} is reduced by one coherence time. The reason that S_{eff} is smaller for a low number of cooperating sensors is the low spatial diversity gain which cannot be compensated by frequency or time diversity. Note that the mentioned tradeoffs are not visible for $b_{\text{sens}} = 1000$ kHz (light, dash-dotted graph in Figure 8.2(a)). Using such a large frequency diversity, the spatial diversity has no influence on S_{eff} anymore. We can also see that $b_{\text{sens}} = 500$ kHz is a good approximation for the optimal sensing bandwidth, except for very small and very large m . In Figure 8.2(b) we show the same results for the resource reservation approach. The maximum achievable S_{eff} is slightly lower for the resource reservation approach which is due to the influence of the link reconfiguration time on the local sensing time in the resource reservation approach

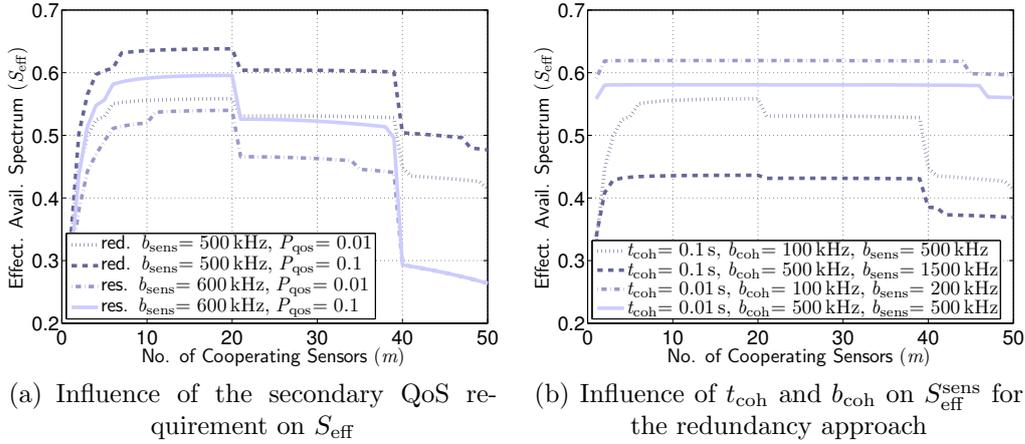


Figure 8.3: Influence of operation environment and QoS requirement on the effectively available spectrum (S_{eff}).

(compare Equation (8.1) with Equation (8.2)). This is also the reason why a slightly increased sensing bandwidth ($b_{\text{sens}} = 600$ kHz) is a better approximation for the optimum sensing bandwidth. Figure 8.2(b) also shows that the resource reservation has a much higher performance loss for high m . The additional reduction of the local sensing time by the distributed sensing time results in a even bigger reduction of the time diversity gain compared to the redundancy approach.

In Figure 8.3 we show the influence of the secondary QoS requirement and the coherence time and coherence bandwidth on the effectively available spectrum. We show results for the sensing bandwidth which best approximates the optimal sensing bandwidth (see Figure 8.2). The influence of the secondary QoS requirement is quantified by the threshold for the probability of violating the QoS requirement of the SUL (P_{qos}). Having a higher threshold leads to less spectral overhead required for the secondary QoS maintenance, which is shown in Figure 8.3(a). For both the redundancy and the resource reservation approach, S_{eff} is increased for a higher P_{qos} threshold. In Figure 8.3(b) we show the influence of the coherence time and coherence bandwidth on S_{eff} for the redundancy approach. We again choose the sensing bandwidth which best approximates the optimal sensing bandwidth. The figure shows that, as expected, the scenario with the highest diversity (light, dash-dotted line) achieves the best S_{eff} and actually needs the smallest sensing bandwidth. For such high diversity even the tradeoff between the spectrum sensing and the QoS maintenance overhead vanishes, due to the very low probability of false positives achieved by the sensing process.

Looking at the general trend of all graphs shown in Figure 8.2 and Figure 8.3 we can see that our proposed system design satisfies the scalability requirement. Even choosing a sensing bandwidth significantly differing from the optimum for small m (which corresponds to initial CR deployments of small SUGs) the CRs can recover some spectrum for secondary usage. As m increases, so does the effectively available spectrum. Remem-

ber that we show only results for simple power measurement-based spectrum sensing. Considering the high diversity scenarios in Figure 8.3(b), even if spatial diversity cannot increase S_{eff} significantly, more sophisticated sensing approaches could be added if the available spectrum does not satisfy the communication needs of the CR network anymore.

8.3 Spectral Efficiency in Dynamic Frequency Hopping Systems

As in the secondary QoS analysis in Chapter 7, we also provide some results and intuition for the performance of Dynamic Frequency Hopping (DFH) systems with respect to the required spectral overhead. Note that the redundancy and resource reservation approaches are specifically designed to work well with non-contiguous SULs. A direct comparison of the partial sensing approach which uses non-contiguous SULs and the DFH approach which uses contiguous SULs would, thus, be unfair. However, we show through our investigation the general applicability of the developed concepts also to other systems and show that the tradeoff between spectrum sensing spectral overhead and QoS maintenance spectral overhead is not constrained to our system design.

8.3.1 System Model

The system model and parameterization in the DFH case is essentially the same as presented in Section 8.2.1 for the partial sensing approach. Instead of partial sensing, however, DFH is used. Thus, a SUL consists of one large sub-channel containing all sub-carriers of the SUL. Considering the example in Figure 7.9 of Section 7.3, this would be one sub-channel consisting of three sub-carriers ($N_{\text{sc}} = 3$). Within one PU band, the DFH approach further alternates between data transmission and spectrum sensing. During the local sensing period, the whole SUL bandwidth is used for the sensing process, i.e., $b_{\text{sens}} = b_{\text{pu}}$, as can be seen in Figure 7.9. Remember that continuous usage of the SUL is possible nonetheless, since during the sensing process within one PU band, the data transmission is continued within another PU band.

The redundancy and resource reservation approaches in case of DFH are very similar as those for partial sensing described in Section 6.4 and shown in Figures 6.5 and 6.6. The main difference is the granularity at which redundancy can be added or resources can be reserved, respectively. In the partial sensing case, a single sub-channel is usually only a small fraction of the SUL bandwidth, and since redundancy is added / resources are reserved on a sub-channel granularity, the number of redundant / reserved sub-channels of a SUL (X) allows fine-grained control of the spectral overhead used for secondary QoS maintenance. In the DFH approach, the whole SUL only consists of one large sub-channel, and consequently the spectral overhead used in case of the DFH approach can only be a multiple of the whole SUL bandwidth (b_{sul}) and does not allow fine-grained control as in the partial sensing approach.

Notice that in case of DFH, the local sensing time (t_{sens}) can be varied as compared to the partial sensing case where the sensing bandwidth (b_{sens}) is varied. The local sensing time is constrained by $t_{\text{sens}} \leq t_{\text{slot}} - t_{\text{coop}}$ for the redundancy approach and by $t_{\text{sens}} \leq t_{\text{slot}} - t_{\text{coop}} - t_{\text{reconf}}$ in the resource reservation approach. The local sensing spectral overhead in case of DFH is not only determined by t_{sens} , but also by t_{coop} and in the resource reservation approach additionally by t_{reconf} . This means that after the local sensing period the PU band cannot be used for data transmission until the end of the slot, i.e., until the distributed sensing (and for the resource reservation approach also the link reconfiguration) is completed.

8.3.2 Performance Analysis

Since the spectrum sensing spectral overhead in case of the DFH approach is dependent on t_{sens} and not b_{sens} as in the partial sensing approach, Equation (8.5) defining the effectively available spectrum after spectrum sensing ($S_{\text{eff}}^{\text{sens}}$) has to be adapted accordingly. For the redundancy approach $S_{\text{eff}}^{\text{sens}}$ can be computed as

$$S_{\text{eff}}^{\text{sens}} = S_{\text{sens}} \left(1 - \frac{t_{\text{sens}} + t_{\text{coop}}}{t_{\text{slot}}} \right). \quad (8.9)$$

For the resource reservation approach $S_{\text{eff}}^{\text{sens}}$ can be computed as

$$S_{\text{eff}}^{\text{sens}} = S_{\text{sens}} \left(1 - \frac{t_{\text{sens}} + t_{\text{coop}} + t_{\text{reconf}}}{t_{\text{slot}}} \right). \quad (8.10)$$

The probability that a PU is detected within its PU band (P_{detect}) is calculated as shown in Equation (8.6). For DFH the whole SUL is placed within one PU band, such that the probability of violating the QoS requirement of the SUL (P_{qos}) simplifies to

$$P_{\text{qos}} = P_{\text{detect}}^{(1+X)} (1 - P_{\text{detect}}), \quad (8.11)$$

with X being the number of redundant / reserved sub-channels of a SUL. The optimum number of redundant / reserved sub-channels of a SUL (X_{opt}) needed to achieve a certain target probability of violating the QoS requirement of the SUL (P_{qos}) can be numerically calculated using Equation (8.11) and the effectively available spectrum can be computed as

$$S_{\text{eff}} = S_{\text{eff}}^{\text{sens}} \left(1 - \frac{X_{\text{opt}}}{1 + X_{\text{opt}}} \right). \quad (8.12)$$

8.3.3 Performance Results

We evaluate the effectively available spectrum (S_{eff}) for DFH using the same parameterization as in the previous section for the partial sensing approach. The most important parameters are summarized in Table 8.3. In Figure 8.4 we show results for S_{eff} in DFH systems for both the redundancy and the resource reservation approach. In accordance with the plots of the previous section, we show results for the local sensing time which

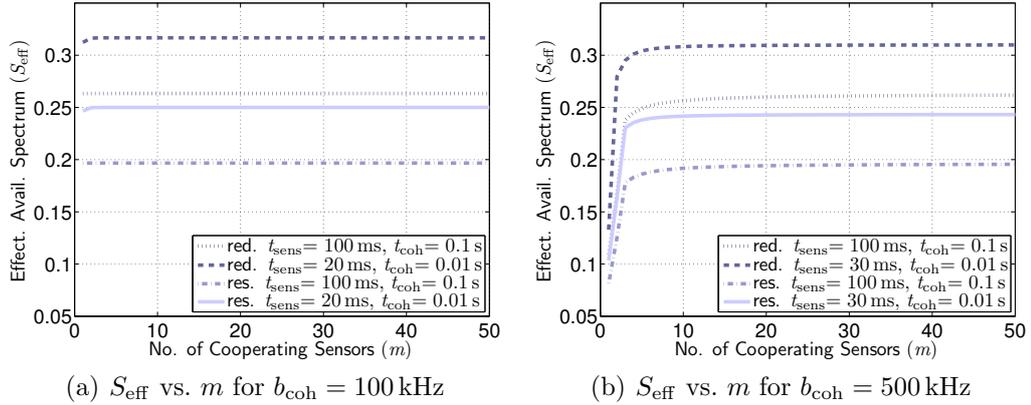


Figure 8.4: The effectively available spectrum for the DFH approach.

best approximates the optimal local sensing time. A first, general observation is the very low effectively available spectrum for all shown graphs. Remember that the granularity at which redundancy is added / resources are reserved is not as fine grained as in the partial sensing approach. By adding one redundant / reserved sub-channel the SUL bandwidth (b_{sul}) is doubled and the effectively available spectrum, thus, reduced to 0.5 at most.

In Figure 8.4(a) we show the effectively available spectrum for a coherence bandwidth of $b_{\text{coh}} = 100$ kHz for the resource reservation and the redundancy approach. Note that for this figure, the shown local sensing times are actually the optimal local sensing times. For all graphs there is almost no tradeoff between the spectrum sensing and the QoS maintenance spectral overhead. The reason is that there is a massive frequency diversity of $M_b = 50$ sensing samples per time slot which basically eliminates all false positives. For the lower frequency diversity case of $b_{\text{coh}} = 500$ kHz, shown in Figure 8.4(b), the expected tradeoff can be observed again. Note however, that the effectively available spectrum does not drop for higher m as in the partial sensing approach. The reason for this becomes clear by looking at the local sensing time: t_{sens} is low enough such that it is not constrained by the distributed sensing time for the values of m plotted in the figure. Figure 8.4(b) also nicely shows the increase of S_{eff} for an increasing m , i.e., the scalability of the spectral efficiency as the network size increases.

This short evaluation of the DFH approach shows that the observed tradeoff between spectrum sensing and QoS maintenance spectral overhead is not only observed for our specific system design, but has to be generally considered when designing CR systems.

8.4 Related Work

The investigation of the spectral overhead required for spectrum sensing has only very recently attracted attention in the research community. Specifically, most of the research in this area tries to optimize the time used for the sensing process with the objective

to maximize the efficiency of the spectrum usage. In [153], the authors propose an optimization approach which finds the optimal sensing time to maximize the channel efficiency, i.e. the fraction of time an idle channel can be detected and used for SU communication. In addition to the sensing overhead, also the energy cost of sensing is considered in [174]. A comparison of energy and feature detection with the objective to minimize the sensing time is presented in [90]. An approach to optimize the joint QoS of the secondary and the primary network based on the bandwidth-time product used for sensing is presented in [54]. The authors of [72] present an approach of optimizing the secondary throughput also considering receiver feedback. There are also some approaches considering the influence of distributed sensing for minimization of the spectral overhead [29, 97, 99, 121]. Note that none of the above publications explicitly considers the frequency domain for the spectrum sensing spectral overhead. It is always implicitly assumed that the whole PU band is used for the sensing process. While the approach to investigate the spectrum sensing spectral overhead taken in this thesis is largely based on [99], we extend the approach to be able to quantify the overhead separately in the time and frequency domain.

While there are investigations on QoS support for secondary communication based on redundancy (e.g. [167]) and resource reservation (e.g. [142]), the spectral overhead required has — to our knowledge — not been investigated yet. Furthermore, we are the first to investigate the *tradeoff* between the spectrum sensing and QoS support spectral overhead.

8.5 Conclusions

In this chapter we have investigated the **spectral efficiency of secondary usage** for a CR system based on the design presented in Chapter 6. We have quantified the spectral overhead required for spectrum sensing and the spectral overhead required for secondary QoS maintenance and have presented a performance analysis investigating the tradeoff between both overheads. From the presented investigation we can conclude that there is a clear tradeoff resulting in an optimal spectrum sensing spectral overhead for which the secondary QoS maintenance overhead is minimized and, consequently, the effectively available spectrum (S_{eff}) is maximized.

Furthermore, we have shown in our investigation that even if the effectively available spectrum is low in the case of very few cooperating sensors (m), the effectively available spectrum increases as m increases. Whereas the qualitative results are the same for the redundancy and the resource reservation approach, the quantitative performance of the resource reservation approach is slightly lower compared to the redundancy approach. This is due to the link reconfiguration time, which reduces the local sensing time in case of the resource reservation approach.

The two main conclusions for the CR system design proposed in this thesis are:

1. Even for very small CR networks with only a few spectrum sensors the effectively available spectrum achievable if applying our system design is significantly greater

than zero. Thus, initial CR deployments with very few, low complexity CRs are possible.

2. The effectively available spectrum increases as the number of cooperating sensors increases. Since the spectrum demand is usually proportional to the number of users, our proposed system design is scalable with respect to the effectively available spectrum it recovers.

CHAPTER 9

Conclusions and Outlook

9.1 Conclusions

In this thesis, we propose and evaluate a Cognitive Radio (CR) system design for opportunistic spectrum sharing which enables low cost, low complexity CR networks to operate in environments with highly dynamic spectrum availability. The presented system design is evaluated based on three metrics: (1) reliable PU detection, (2) QoS support for the secondary communication, and (3) efficient spectrum usage by the CR network.

CR systems follow the approach of using licensed spectrum, which is temporarily not used by the respective license holder (PU), for their own communication needs. This secondary usage of spectrum is under the premise that the PU has strict access priority to its spectrum. A CR network, thus, has to monitor the spectrum it is using for the appearance of the PU and switch to other temporarily available spectrum if the PU is detected. Consequently, the preferred PU bands for CR networks to operate in are owned by PUs with a rather static behavior. This is one reason why one of the first deployed CR systems will operate in the TV bands.

In this thesis we make the case for considering secondary usage also in environments with a very agile spectrum availability. We show that the variability of available spectrum does not only depend on the usage behavior (agility) of the PUs but also on other factors which can significantly increase the variability of available spectrum *perceived* by the CR system. The first factor to consider is the mobility between the PU transmitter and the CR system: if the mobility of the PU or CR system is high in comparison to the transmission range of the PU, the CR system will experience a high variability of the available spectrum even if the PU is continuously using its spectrum. The second factor influencing the variability of available spectrum is the accuracy of PU detection. For the most flexible secondary usage scenario of opportunistic spectrum sharing considered in this thesis, PU detection is done through a sensing process. In order to assure reliable PU detection, the sensing process might be dimensioned to be “oversensitive”, i.e., declare the detection of a PU although no PU has appeared. These false positives result in an artificially increased variability of the available spectrum: the CR system assumes the spectrum to be occupied, no matter whether the detection was caused by a real PU (true positive) or a false positive. Having a high probability of false positives (P_{fp}) can, thus, significantly increase the variability of available spectrum. Following this line of argument, high variability of available spectrum is a real and valid use case for secondary spectrum usage.

We show that reliable detection of PUs, quantified by the probability of false negatives (P_{fn}), is possible also with low accuracy, simple energy detection based spectrum sensors. The price is a potentially very high increase of the probability of false positives (P_{fp}). We further show that, given a fixed, very low P_{fn} , P_{fp} can be reduced by increasing the spectral overhead used for the sensing process and by using diversity approaches for spectrum sensing. We propose a CR system design using such a simple, energy detection based sensing process. A partial sensing process is proposed, which uses only a part of the PU band for spectrum sensing such that the rest can be used for data transmission in parallel, which eliminates the need to interrupt the data transmis-

sion to do spectrum sensing. We further use the concept of non-contiguous SULs for the proposed CR design. The secondary communication link (SUL) is split into several small sub-channels which are all located in different PU bands, such that, if a PU appears, only a small part of the SUL has to be reconfigured. These two concepts of partial sensing and non-contiguous SULs greatly improve the QoS of the secondary communication. We show that a large number of sub-channels for an SUL usually achieves a superior QoS compared to a small number of sub-channels assuming that the total available bandwidth is equal. We further apply redundancy and resource reservation approaches to our CR system, which further improve the secondary QoS. The improvement, however, comes at the cost of spectral overhead, decreasing the efficiency of our CR system design.

Through the evaluation of this proposed CR system design we demonstrate that it is in fact possible to start deployment of CR systems with very simple and cheap sensing accepting high false positives but assuring proper QoS for the primary **and** secondary communication. Admittedly, the spectral efficiency of the secondary usage will suffer, but given the abundance of unused spectrum, the available capacity for such modus of operation will — at least initially — be sufficient. Thus, it might be a reasonable approach to start applying initially rather simple sensing systems (assuring PU protection, i.e., a low probability of false negatives, and accepting a high probability of false positives), and maintaining the secondary QoS by using a proper (very high) amount of spectral overhead. As the number of CR systems / the traffic carried by the CR systems increases, the quality of the sensing process can be increased resulting in a better spectral efficiency. We show in this thesis that following such a concept is possible with the proposed CR design. As the accuracy of the sensing process increases (i.e., by an increase of CRs participating in the sensing process) also the spectral efficiency increases. This makes the proposed CR design an ideal approach for initial deployments of very small and cheap CR networks, which are perfectly scalable: with an increase of network size, also the spectral efficiency is increased.

9.2 Future Work

In this thesis, we have presented and evaluated a general CR system design which enables low cost, low complexity CR networks to operate in environments with highly dynamic spectrum availability. While we have evaluated and validated the main concepts and most important building blocks for such a CR system, there are still several aspects that should be investigated.

For the cooperative sensing as well as for the secondary QoS maintenance process signaling protocols are required such that the CRs can achieve a joint sensing result and decide on which sub-channels to use for their communication. For those signaling protocols a control channel is required, which might be a dedicated physical channel or a logical channel. For CR systems, where spectrum for communication is usually only available on an opportunistic basis, the implementation of such a control channel is a challenging task. Although control channels for CR and signaling protocols for distributed sensing and secondary link and QoS maintenance have been considered in

the research community for quite some time, the influence of such protocols on the presented CR system design needs to be investigated. Specifically, the influence of the protocol spectral overhead of distributed sensing and secondary QoS maintenance on the performance results presented in this thesis are an interesting aspect for future work.

Another aspect for future work is the prototype implementation and experimental evaluation of the proposed CR design. The results presented in this thesis are mainly based on analytical investigations and simulations. While those are important tools especially for early system design considerations, eventually at least selected functionalities should be experimentally validated. It would, e.g., be interesting to see if the analytical results for the diversity approaches for spectrum sensing can be reproduced in real systems.

APPENDIX **A**

Modeling Cellular Primary Usage

In this chapter we present a model for the PU agility of Primary User operating in the cellular bands. The presented model is based on the measurement campaign of Section 4.3 and has been published in [170, 171].

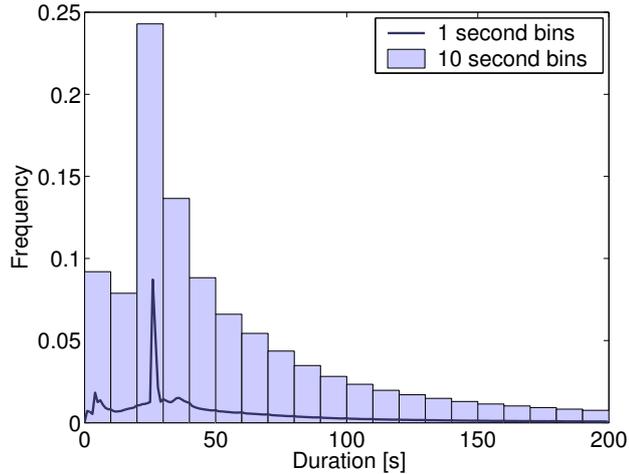


Figure A.1: Histogram of call durations. We plot the histogram using different bin sizes.

Since secondary users opportunistically use spectrum not utilized by primary users, models of primary usage in individual cells play an important role in designing and deploying cellular DSA approaches. Based on the measurement campaign of Section 4.3 we developed models for the PU agility of cellular PUs. There are two simple models that fit the behavior of primary users well; a call-based and an event-based model.

A.1 Call-based Model

The *call-based model* uses two random variables, T and D to understand the dynamics of spectrum usage. T is the random variable corresponding to the inter-arrival time between two calls and D corresponds to the duration of calls. Under this model, the system can be described using only T and D . An obvious and popular choice is to model call arrivals as a Poisson process (i.i.d. with exponential inter-arrival times) and call durations as being exponentially distributed.

It turns out that the distribution of call inter-arrival times is well described by an exponential distribution in more than 90% of the hours for most cells. We use the *Anderson-Darling* test with 95% confidence level as a goodness-of-fit test for exponential distribution. Note that, since we use a 95% confidence level for the Anderson-Darling test, we expect only 95% of our tests to succeed. The technical details of these tests can be found in [170].

We also calculate the auto-correlation coefficient for each per-cell per-hour sequence of call inter-arrival times. We find that only 20% of these sequences have auto-correlation coefficients (at non-zero lags) higher than 0.16. Though not conclusive, such low auto-correlation is consistent with independence. Hence, we believe that call inter-arrivals are well-modeled as an exponentially distributed i.i.d. sequence. In other words, call arrivals can be viewed as Poisson processes. Though Poisson processes have been used to model

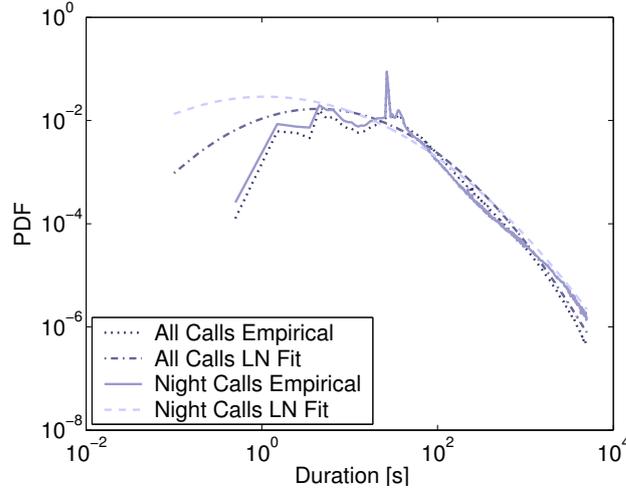


Figure A.2: Duration distributions and lognormal fits.

fixed telephone calls for a long time, our study is one of the first to show that this is largely true for *individual cells* in mobile systems.

Though the inter-arrival times of calls are well modeled as a Poisson process, the call-based model has a significant disadvantage, namely, the distribution of call durations. In Figure A.1, we plot the empirically-observed histogram of call durations. The histogram is quite unlike that of an exponential distribution. In fact, the histogram is not even monotonic. We see about 10% of calls having a duration of about 27 seconds. These correspond to calls during which the called mobile users did not answer and the calls were redirected to voice-mail. However, RF voice channels were allocated during these calls. This illustrates that call durations can be significantly skewed towards smaller durations due to non-technical failures, e.g., failure to answer. Also, note that the variance of the call durations is more than three times the mean, which is significantly higher than that of exponential distributions.

Further analysis shows that there are likely two different distributions of call durations — one during the day and the other during the night (11PM-5AM). Furthermore, the transition hours between day and night likely see a mixture of both these distributions. In Figure A.2, we compare the overall and nighttime distributions of call durations. Note that we use the log-log scale. We find that the nighttime distribution has more short calls as well as a heavier tail compared to the overall distribution. Both distributions have a “semi-heavy” tail and are not well-modeled by classic short-tailed distributions such as Erlang (results not shown). However, the shape of the above distributions is reminiscent of the lognormal distribution, which is parabolic in log-log scale. Recall that D is lognormally distributed with parameters μ and σ^2 if $\log(D)$ is normally distributed with the same parameters. In Figure A.2, we also plot the best lognormal fits for the distributions of call durations. The head of the empirical distribution shows significant deviation from the best lognormal fit. Although the tails of the empirical and

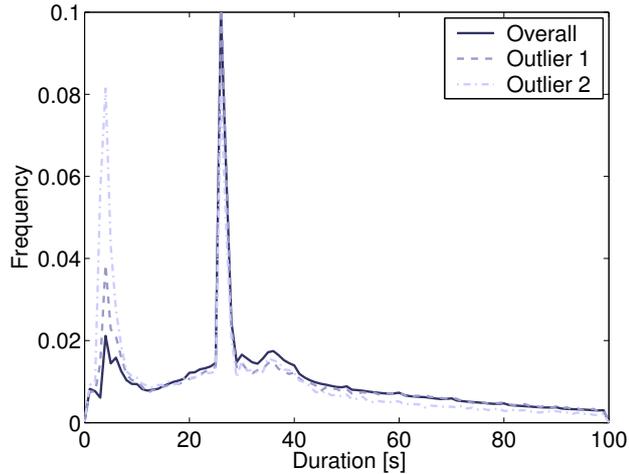


Figure A.3: Illustration of anomalous distributions during 2 hours.

best fit agree better, they too diverge at large values.

Not only is the distribution of call durations hard to model, there can also be significant deviations during certain hours. We plot two such “outlier hours” in Figure A.3. The two outlier plots correspond to the weekday hours plotted in Figure 4.3 — the spikes in the arrival rate correspond to the spikes of Figure A.3. Both hours see a sudden spurt in short calls. We verified that at least one of these is caused by a large number of calls to a popular television show, whose telephone lines are often busy. Figure A.3 thus demonstrates that social behavior and external events, which may not be easily predicted, can and do have significant short-term impact on spectrum usage.

A.2 Event-based Model

The skewed distribution of call durations is the primary disadvantage of the call-based model and can be eliminated by using an alternative *event-based* model. This model ignores details about individual calls and instead models only the load $X(\cdot)$, i.e., the total number of ongoing calls. Under this model, the load is considered to be a one-dimensional, continuous-time random walk where steps are either $+1$ or -1 , corresponding to the initiation and termination events of a call:

$$X(t + E) = X(t) + (-1)^\Phi. \quad (\text{A.1})$$

Here E is a random variable representing the time between consecutive steps/events and Φ is a Bernoulli random variable, which takes the value $+1$ with probability p and 0 otherwise. Since there is a $+1$ for every -1 , p should be $1/2$. A Poisson process is the obvious choice to model the occurrence of events resulting in exponentially distributed inter-event times.

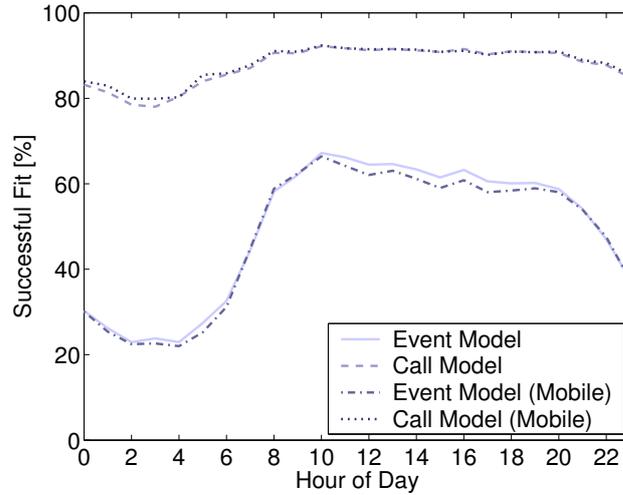


Figure A.4: The percentage of successful fits (across all cells) averaged on a per-day basis.

It turns out that inter-event times are well-modeled as exponential distributions for only about 50 % of the hours in most cells. As shown in Figure A.4, exponential modeling fails almost twice as often during the night (when the load is low) than during the day.

The skewed distribution of call durations is also responsible for the failures of the event-based model. This is because the $+1$ and -1 events correspond to the initiation and termination of a call and are separated by the duration of that call, which is not exponentially distributed. If there are no additional events during the duration of that call, the duration itself will be an inter-event time. In general, call durations or portions thereof will be part of the inter-event times. Thus, during the hours of the night when the system load is low, the non-exponential distribution of call durations has a significant impact on the distribution of inter-event times. During the day, this impact is reduced.

A second component of the event-based model is that the ± 1 events form a Bernoulli process. A necessary (but not sufficient) condition for this to be true is that the sequence of ± 1 s should have close to zero auto-correlation at non-zero lags. To understand if this is true, we plot the mean autocorrelation at non-zero lag values on an hour-of-day basis in Figure A.5. We see a similar effect as above. During the nighttime, when the load is lower, the $+1$ of a call is more likely to be followed by the -1 of that call. This causes negative correlation at odd lags. Accordingly, we can also see the positive correlation at even lags. During the day this effect is reduced.

The above discussion shows that the event-based model is more applicable when the load is high, though the Bernoulli assumption is not strictly valid. However, when the load is low, the call-based model with a skewed distribution of call durations is the superior model.

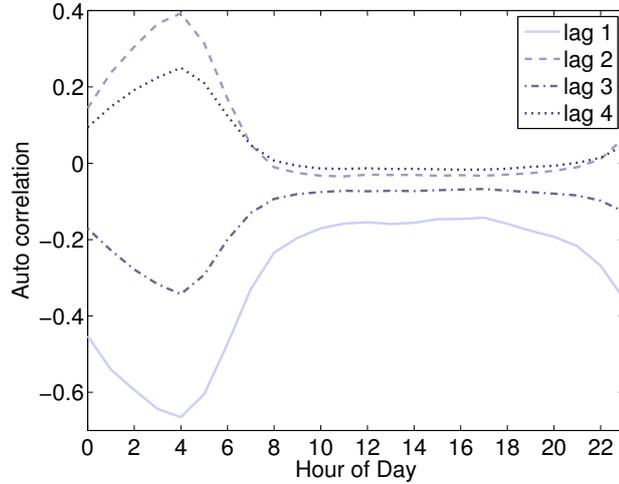


Figure A.5: The per-hour auto-correlation of the step sizes (Φ) in our event-based models averaged across all cells.

A.3 Implication for Spectrum Pricing

Characterization and modeling of PU spectrum usage provides several insights that are crucial to enable secondary usage of spectrum. For example, the owners of spectrum need models of their PUs to determine how much secondary usage is feasible and how it can be priced. Models for call arrival and call duration are essential for optimal pricing strategies of auctioned spectrum. In [5, 118] the authors develop optimal pricing strategies for secondary usage of cellular CDMA networks. The strategies only depend on the call arrival and call duration distributions, which are both assumed to be exponential. Our results show significant deviations of call durations from exponential distributions. Hence, these strategies may have to be revised. The precise implications are subject to further studies.

A.4 Implications for Spectrum Sensing

From a CR perspective, there are two fundamental questions to be answered for the development of sensing techniques for CRs: (1) How often must sensing be performed? (2) What is the required observation time of a single channel to reliably detect potential PUs? Answers to these questions determine how much time and resources are needed for detecting PUs.

The first question is usually answered by the PU, which specifies the maximum interference time (t_{\max}), i.e., the maximum time a SUG is allowed to interfere with a PU communication. Clearly, the maximum interference time sets an upper limit on the periodic time interval after which a channel used by a SUG has to be sensed. Knowing the probability distribution of the arrival process of the primary communication (in our

study the call arrivals), and given a target probability p_i that the SUG interferes with the PU, t_{\max} can be simply calculated using the CDF ($p_i = P(X \leq t_{\max})$). Equation (A.2) shows the calculation of t_{\max} assuming an exponential call arrival process.

$$p_i = 1 - e^{-\lambda t_{\max}} \Leftrightarrow t_{\max} = -\frac{\ln(1 - p_i)}{\lambda} \quad (\text{A.2})$$

The knowledge of the arrival process, thus, enables us to adjust the maximum interference time. For our investigation, the mean call inter-arrival time (over one hour) per cell varies from the sub-second range to tens of minutes. Assuming a maximum of 30 calls per cell and a probability of interference of $p_i = 0.001$ this would result in a required maximum interference time between $t_{\max} = 0.03$ s and $t_{\max} = 18$ s. This huge gap clearly indicates the gains achievable by choosing t_{\max} based on the call inter-arrival time, which can itself be gleaned by sensing. Results such as those in Figures 4.6 and 4.7 also provide insights into good tradeoffs for sensing strategies.

An answer to the second question, i.e., determining the time needed for sensing a *single* channel, is much more complex and depends on various factors such as the sensitivity requirements of the PU, the specific sensing technique used, distributed/cooperative sensing aspects, etc. However, regardless of the time the sensing process takes for a specific system, it is desirable not to waste this time for sensing an occupied channel. Here, a model of the duration of a PU communication can help to determine the time after which a channel sensed to be occupied by a PU should be sensed again. In particular, our analysis of the call durations shows that there are many short calls and the remaining are spread over a “semi-heavy” tail. Hence, a conditional sensing process is well-motivated: the CR initially uses a rapid sensing frequency for the case that a new call is short. After a few tens of seconds, rapid sensing is likely to yield little benefit. Hence, slower sensing is justified.

A.5 Related Work

In this section, we summarize models developed based on the measurement campaigns presented in Section 4.1. Note that all models discussed below have been published after the model we have developed and presented in this appendix. Furthermore, none of these models provides the fine-grained time resolution of our model.

In [53] a semi-Markov model for the spectrum occupancy in an IEEE 802.11b [83] network is proposed. The study is motivated by the fact that for a Skype conference call, the channel is idle for about 90% of the time. Ghosh et al. [58] proposed a model based on exponentially distributed inter-arrival times and an exponential distribution of the idle times. The model is validated against measurements done in 928 – 948 MHz paging band.

In [158, 159] the authors develop a model for the duty cycle based on the data from the Aachen measurement campaign [135]. The authors calculate the average duty cycle per hour and show that the distribution of the hourly duty cycles is well approximated by geometric and lognormal distributions. The model is validated for various bands /

technologies (cellular, DECT cordless phones, WLAN). In [161], the model for time variability is also extended to model the variability of occupancy in the frequency domain. It is shown, e.g., for the 1800 MHz GSM that whereas some channels seem to be continuously occupied or idle (as already explained above), especially the bands at the edge of highly occupied or vacant bands tend to have a very dynamic occupancy.

Within the Guangdong measurements [24] models for the idle time durations (referred to as channel vacancy duration) and the instantaneous duty cycle (referred to as service congestion rate) are developed. It is shown that the idle time durations follow an exponential-like distribution (in the form of $y = a + be^{-cx}$) but are **not** exponentially distributed. They, thus do not have the memoryless property of the exponential distribution, which means there is significant correlation in time. The advantage of the instantaneous duty cycle over the models used in the above works is that the time variant information is not lost, such that models for the short term variation can be developed. An interesting observation is that the instantaneous duty cycle shows a strong correlation, whereas the idle time durations show no clear pattern. The reason is that the instantaneous duty cycle is averaged for a set of adjacent PU bands offering the same service.

APPENDIX **B**

Reliable Detection of Cellular Primary Users

We have shown in Chapter 5 that sensing-based detection of PUs is — in principle — possible, even only using few sensors and applying simple energy detection. The main enabling concept is to accept a rather high probability of false positives (P_{fp}). This chapter presents an experimental evaluation of sensing-based detection for a very specific class of PUs: CDMA-based cellular telephony. We show that the general concept developed in Chapter 5 can also be applied to such very challenging PUs. A condensed version of the work presented in this chapter has been published in [173].

For TV bands as well as other usage scenarios discussed so far, a SUG needs to reach a binary decision, namely, whether a PU is present or not. In the case of CDMA-based cellular networks, the situation is quite different. Since CDMA voice users share spectrum, the goal is not to reach a binary decision whether a PU is present or not but to identify the amount of *unused spectrum capacity*.

In this chapter, we investigate how unused spectrum capacity can be estimated by utilizing energy-based sensing mechanisms. Not only are these mechanisms simple but they are also cheap (thus enabling large-scale CR deployment). We phrase our investigation in terms of the following problem: **What information about PU occupancy do spectrum measurements of energy yield? Can this information be used to estimate the unused capacity that SUGs can exploit?**

To conduct our investigation, we leverage a large and unique dataset containing spectrum measurements of transmitted power *synchronized* with the “ground truth” about the cellular band usage, which is extracted out of the provider’s call records collected at the network switches. To our knowledge, this is the first study to execute and evaluate such synchronized measurements of sensing and ground-truth.

B.1 Measurement Methodology

To evaluate if sensed power yields information about the underlying primary usage, we conducted a large number of sensing measurements spread over time and multiple locations. In addition, we simultaneously collected detailed information about primary usage as recorded inside the network. Thus, we were able to collect a large amount of unique data consisting of both measurements *and* ground truth. Our measurements were collected in the band used by a CDMA-based cellular operator. In total, our experiments yielded approximately 90 GB of ground truth data on primary usage and 14 GB of data with spectrum power measurements. In this section, we describe our measurement methodology in detail and also the salient aspects of the collected datasets.

B.1.1 Ground Truth on Primary Usage

To compute the ground-truth, we use call records collected at the switches of a CDMA-based cellular operator. These records capture the start time, duration, initial and final sector as well as assigned carrier¹ of the voice calls made in the network. The call duration reflects the RF emission time of the call and captures precisely what we want — the duration of primary usage. All timestamps were measured with a resolution of a few milliseconds.

Using the call records, we calculate the ground-truth usage in each carrier of a cell sector. We split the call records based on the sector and carrier. We create two records for each call corresponding to its initiation and termination. Then, we sort these records in order of their time to get a sequence of call *events* that are known to have occurred in

¹Each cell sector may be assigned one or more carriers — a 1.25 MHz portion of the spectrum. Each carrier is capable of supporting tens of calls.

(a) Measurement location $D1$ (b) Measurement location $D2$

Figure B.1: Measurement locations (in red) and nearest cell tower (in blue).

each sector / carrier. Using the sorted list of events, we also calculate the *load* of each sector and carrier. We do so by maintaining a running count of the number of ongoing calls. This count is increased by $+1$ when a call begins and decreased by -1 when a call terminates.

Since the switches only record the initial and final sector of each call, we are unable to account for the spectrum usage in other sectors that the user may have visited in between. This implies that our list of events is accurate but not necessarily complete (it may not count call initiation and termination events caused by handover). Similarly, for calculating load, we assign the whole call to the initial sector / carrier. We also try other approximations of load by, for example, assigning the first half of the call to the initial sector / carrier and the last half to the final sector / carrier. Since these approximations do not alter our results significantly, we do not provide them. Furthermore, we believe that full mobility information is unlikely to change our results either.

B.1.2 Power Sensing

We collected our power measurements using a W1314A Multi-band Wireless Measurement Receiver from Agilent [2]. We used the Model 110 so that we can sense the uplink and downlink CDMA bands (in the 1900 MHz frequency range) used by the network whose call records we had access to. This receiver captured in real-time power measurements from all 12 1.25 MHz carriers belonging to a single provider. The power measurements were reported twice or thrice a second on average. Unless stated otherwise, we convert this raw data into per-second averages (computed using 2 or 3 readings). Our wireless measurement receiver is a sophisticated piece of equipment and can be viewed as being capable of the *most accurate* power-based sensing that a CR can perform.

We collected multiple datasets at four different urban locations ($L1$ to $L4$). For each of the locations the power measurements were collected over multiple days and spanned all possible 1.25 MHz carriers of the CDMA band. We use $C1, C2, \dots$ to refer to the various carriers. $L1$ was within line-of-sight and about 0.5 miles from the nearest base station. $L2$ was at a similar distance but not within line-of-sight of the closest base station. In both $L1$ and $L2$ the antenna was placed close to a window on the second floor. $L3$ and $L4$ were about twice the distance (1 mile) to the closest base station, on the ground floor and heavily shadowed from the closest base station.

Our analysis showed the following trends: datasets collected at the same location (at different times) showed similar results. Furthermore, the datasets collected at locations $L1$ and $L2$ illustrated better results than those collected at locations $L3$ and $L4$. This was due to the heavy shadowing at the latter locations. On account of this and space limitations, we provide results only for $L1$ and $L2$. In particular, we use two datasets, which we label $D1$ and $D2$. $D1$ was collected over a period of 3 days at location $L1$ and $D2$ was collected over a period of 4 days at location $L2$. We present results for two carriers $C1$ and $C2$, which were the most active carriers in the two locations. In Figure B.1, we show the location at which we collected our measurements.

For each experiment, we also recorded the identities of the cell sectors with the strongest pilots. For each dataset, we refer to the cell sectors as $S1, S2, \dots$ in decreasing order of pilot strength. We observe that, for both of our datasets, this order reflected the distance of the sectors from the measurement locations. Usually a cell sector has activity in more than one carrier. We denote this by referring to the activity of sector 1 in carrier 1 with $S1.1$, in carrier 2 with $S1.2$, etc. Note that, for each of these cell sectors, we computed ground-truth (events and load) using the call records as described in Section B.1.1.

B.2 Estimation via Power Thresholds

In this section, we present how well power is correlated with primary usage. We obtain these results by investigating the following question: Can a simple scheme based on mapping sensed power information to load information work? We start with some preliminary observations and data analysis for synchronizing the power measurements and ground truth.

B.2.1 Power-Load Correlation

In Figure B.2, we plot the power sensed and in Figure B.3 the network ground truth for dataset $D1$ (downlink). We plot results for the two strongest sectors $S1$ and $S2$ and two carriers $C1$ and $C2$ used by both sectors. For data confidentiality reasons, we normalize the load values by a randomly chosen number so that the absolute load is obfuscated while preserving the trends. Notice that the day / night variations of the load are clearly visible in Figure B.3. This is not surprising since it corresponds to levels of human activity and has also been observed in prior work [89, 170]. Figure B.2,

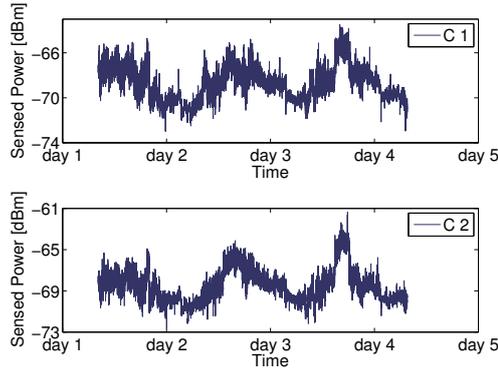
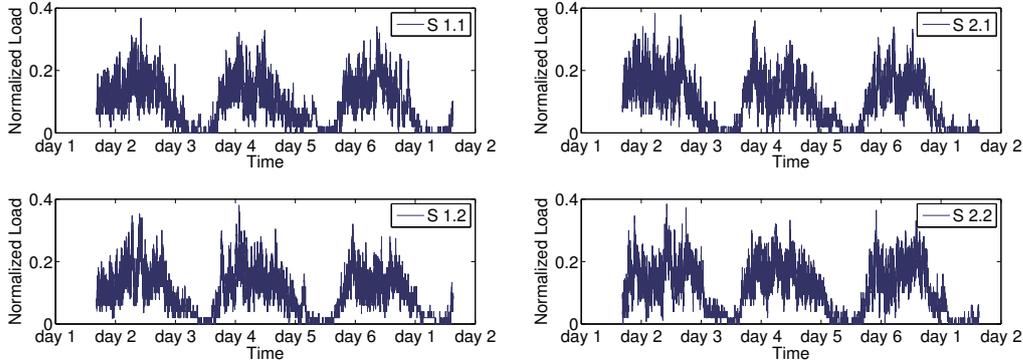


Figure B.2: Dataset $D1$: Sensed power over time for carriers $C1$ and $C2$.



(a) Strongest cell sector $S1$.

(b) Second strongest cell sector $S2$.

Figure B.3: Dataset $D1$: Normalized load for strongest cell sectors.

showing the sensed power, also illustrate a distinct diurnal pattern with higher power levels during the day and lower levels at night.

Cross-correlation is a simple and well-known metric that we can use to measure the extent to which power “tracks” load. We plot the cross-correlation between sensed power and load for several lags in Figure B.4(a). The maximum is not reached at lag 0 since the clocks used for collecting call records and power measurements are only synchronized to within a few seconds of each other. In addition, for each sector / carrier, we see peaks separated by roughly one day. These local maxima are caused by the underlying diurnal pattern of the load. Overall, it is clear that the sensed power is indeed well correlated with load.

It turns out that the cross-correlation curves for all sector / carriers reach a maximum at the same lag of 13 seconds. This is shown in the bottom plots of Figure B.4(a). We observed maxima at a similar lag with dataset $D2$ as well. The remarkable consistency of the lag at which the maximum correlation is attained allows us to assert that this is

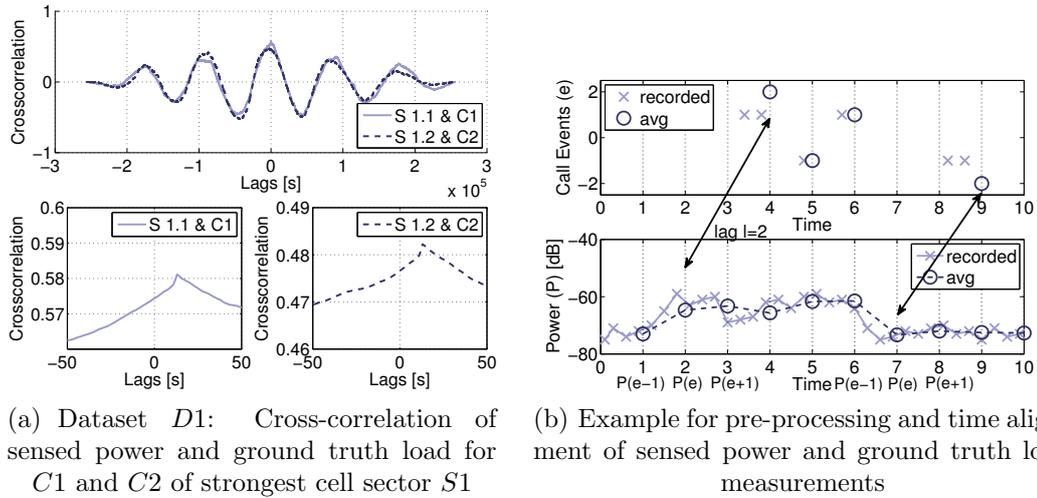


Figure B.4: Cross-correlation and time synchronization of sensed power and ground truth load

indeed the difference between the clocks used for call records and power measurements². For the rest of this work, we use the above lag to synchronize the power measurements and call records.

For clarity, we illustrate the synchronization process between the raw power measurements and call records in Figure B.4(b). Recall that we obtain 2 – 3 sensed power measurement readings of each carrier every second. These are averaged (by default) into per-second measurements. Similarly, the event / load information is aligned to a grid with integral number of seconds. The averaged power readings and grid-aligned call records are then synchronized as shown in Figure B.4(b). The synchronization is based on the lag 13 obtained from the cross-correlation curves.

B.2.2 Naive Threshold-based Scheme

We start investigating if there is a unique mapping of power level to sector load in Figure B.5(a). In this figure, we plot the distribution of the sensed power levels for various coarse-grained levels of load during peak hours of the day (12 p.m. to 6 p.m.). As expected, with increasing load, the power levels tend to increase. But, observe that for different values of load, the same power levels are often seen. Moreover, when the load is low, the power levels are more spread out. This indicates that it is challenging to distinguish small changes in load using static power alone.

However, power levels are better able to separate the coarse-grained measures of load at night (12 a.m. to 6 a.m.). We show this in Figure B.5(b). Note that, since the load at night is never high, we only show two levels. These levels are relatively well separated: for example, the power level is below -70 dBm for 60 % of the time when the load is low

²The lag that maximizes correlation has often been used to synchronize clocks.

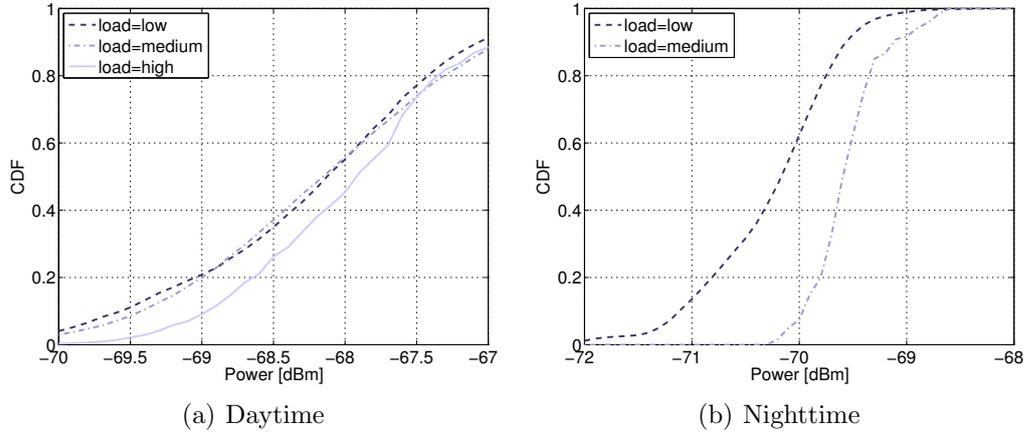


Figure B.5: Dataset *D1*: Distribution of power levels when the load varies.

as opposed to 5% of the time when we have medium load. Such power-based thresholds separating coarse-grained load levels such as low, medium and high are likely to depend on location. For example, in separate short experiments, we found that the average power levels close (about 100 m) to a base station were around -50 dBm. Though we did not find much variation at a given location within a few days, it is unclear how long thresholds at a certain location remain valid.

To summarize, load estimation based on power thresholds (that are gleaned on a per-location basis) can provide coarse-grained information about load, especially at night. Though fine-grained information is difficult to extract, such coarse-grained information might be sufficient to decide whether or not to start secondary usage. We note that the above power threshold-based scheme is a *black-box* estimation of load in the sense that no information is required from the cellular operator.

B.3 Event Signatures

During our analysis, we found that sensed power contains information about the “first derivative”, i.e., change of primary usage load. These changes are typically due to call initiation and termination in a CDMA voice network. We consider the downlink and uplink in that order.

B.3.1 Correlating Sensed Power with Call Events

Figure B.6 provides a simple illustration of the potential behind event detection. In this plot, we show a typical example of measured power over a couple of minutes. We also plot the call initiations and terminations at the strongest sector *S1* during this time period. In several cases of call initiation, the power increases, and for several cases of call termination, the power decreases. For example, we see a large decrease in power labeled “c” when a call is terminated at *S1*. However, we also see counter-intuitive behavior of

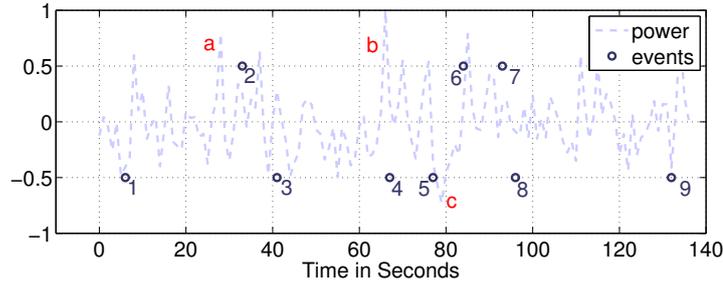


Figure B.6: An illustration of how power changes when various call events happen. The higher row of circles indicate call initiation events and the lower row indicate call termination events.

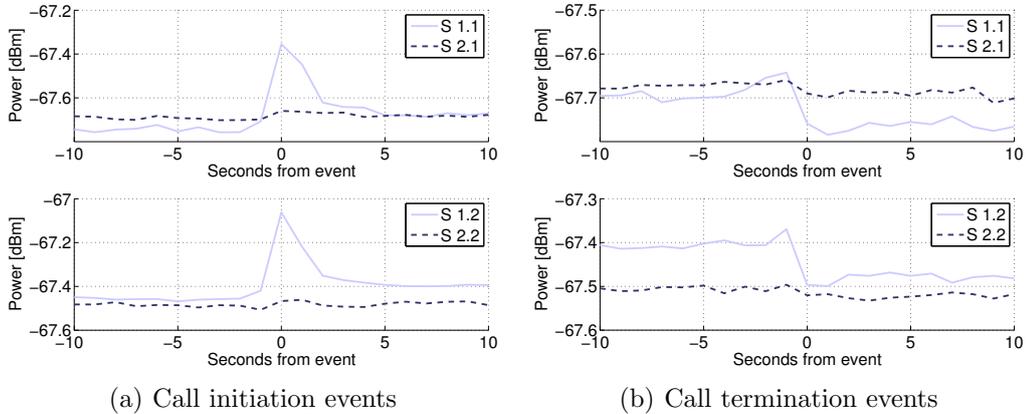


Figure B.7: Dataset $D1$: 10-second t -average plots for (downlink) power during call initiation and termination events. The upper pair of plots are for carrier $C1$ and the lower pair for $C2$.

sensed power. For example, observe the large increase labeled “a” without any event and the increase labeled “b” after a termination event. Also, we see significant power variations even when no events occur. This is the case between the first and second event in Figure B.6 when the sensed power varies over a range of more than 1 dBm.

Thus, there does appear to be significant potential to detecting events by looking at power changes — a more complex approach to power thresholds. The main challenge seems to be that of minimizing false detection.

B.3.2 t -average Plots

As discussed above, the sensed power in our datasets often shows jumps and drops when events occur. To understand if such event *signatures* exist and characterize them, we examine all the initiation and termination events spread across multiple days in our

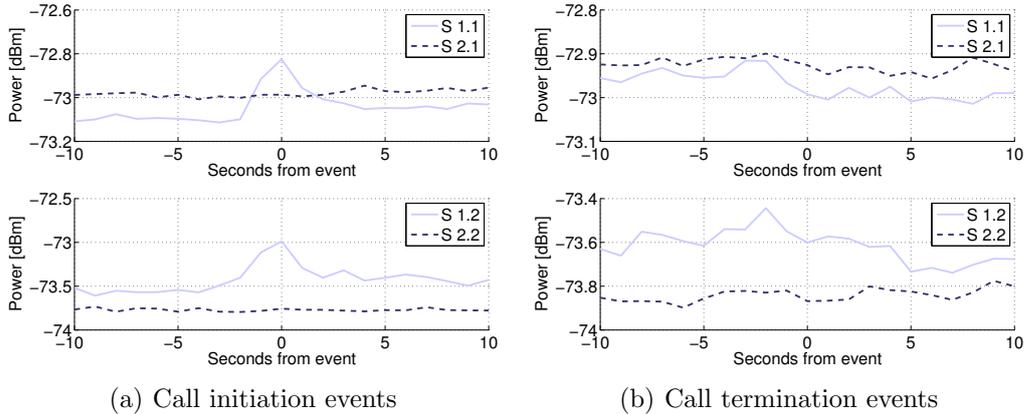


Figure B.8: Dataset *D2*: 10-second *t-average* plots for (downlink) power during call initiation and termination events. The upper pair of plots are for carrier *C1* and the lower pair for *C2*.

datasets. Since the average power across these days may not be stationary we rely on what we refer to as *t-average* plots: we extract the power for a short time periods before *and* after events. Then, we average the sensed power across all initiation and termination events. This allows us to look at time-averaged characteristics of power when events occur even in the absence of stationarity — a key advantage.

In Figure B.7, we show a *t-average* plot capturing the average behavior for 10 seconds before and after call initiation and termination events of dataset *D1*. In this figure, we are considering only the downlink bands. We notice two key patterns for the events involving the strongest cell sector *S1*:

1. On average, call initiation events are characterized by a spike that is about 0.3 – 0.4 dBm followed by a general increase of power afterwards that is around 0.05 – 0.1 dBm. This trend is clearly seen during the daytime and nighttime as well (not shown). We believe that this spike is easily explained by the CDMA downlink power control loop [151]. This loop ensures that, when a call starts, the base station transmits with high power. The power level is then reduced to a minimal level while maintaining call quality using the rapid closed-loop power control of CDMA. The increase of 0.05 – 0.1 dBm reflects the increased power due to a new call, of course.
2. On average, call termination events are characterized by a general dip of at least 0.05 – 0.1 dBm (a bit higher during nighttime) immediately after the event. This reflects our intuition that the call termination corresponds to lesser power being emitted by the base station (we are looking at the downlink power).

Thus, there exist well-defined power signatures for initiation and termination events in CDMA networks. Figure B.7 also shows that there are no visible signatures corresponding to the events of the second strongest cell sector *S2*.

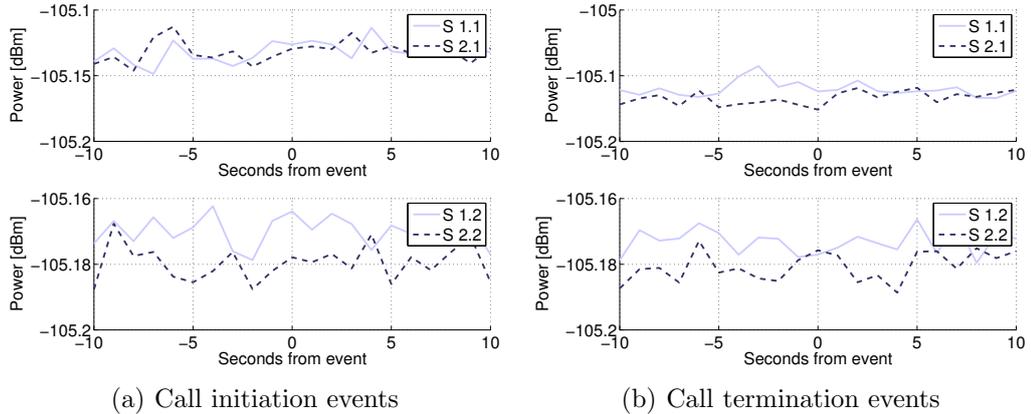


Figure B.9: Dataset $D1$: 10-second t -average plots for (uplink) call initiation and termination events. The upper pair of plots are for carrier $C1$ and the lower pair for $C2$.

Recall from Section B.1.2 that we collected a second dataset $D2$ from experiments conducted at another location. To verify that the event signatures persist across locations, we plot t -average plots for $D2$ in Figure B.8. The call initiation signature corresponding to events of the strongest sector $S1$ continues to be clearly seen for both carriers. As before, there is no signature corresponding to events of the second strongest sector $S2$.

Surprisingly, Figure B.8 shows that the call termination signature for $D2$ is less clear. Recall from Section B.1.2 that the location at which dataset $D2$ was collected had no line-of-sight to $S1$ or $S2$. We believe that this is the likely reason behind the weak termination signature.

B.3.3 Uplink Events

We have seen that downlink power measurements exhibit well-defined (average-case) signatures for call initiation events and, to a lesser extent, for call termination events. In Figure B.9, we use t -average plots to investigate if such signatures are also present in the sensed power of uplink bands. We find no identifiable signatures corresponding to initiation or termination events. This is true for events of $S1$ and $S2$.

The absence of signatures in uplink power is not surprising given that the average power levels are about 25 – 30 dBm lower than the downlink power measurements. Such low power levels are likely due to the stricter power budget of end-user devices. Also, these end-user devices are spatially distributed and we expect to see signatures when such devices are nearby. To verify this hypothesis, we conducted “active” experiments by initiating phone calls using a mobile handset located near our power sensor. When these calls were initiated, we did observe identifiable spikes in power similar to the downlink initiation signature. These experiments confirm that uplink sensing is of use only if the

sensor is close to all end-user devices. Since this is physically impossible, we do not further investigate uplink sensing.

B.4 Event Detection

In the previous section, we found the presence of well-defined signatures corresponding to call initiation and termination events on the downlink CDMA channels. In this section, we show that such average-case signatures do not, however, translate into algorithms for accurate event detection.

B.4.1 Discriminators of Initiation Signature

The *t-average* plots discussed in the previous section indicate that call initiation can potentially be identified by detecting *spikes* in the sensed power. Referring to the *t-average* plots of Figure B.7, we see that there are roughly three time periods: during, before and after call initiation, each of which consists of multiple seconds. The spike occurs during the call initiation and is significantly higher than the power before and after. We also expect the power after call initiation to be higher than before.

Based on the above discussion, we are motivated to consider three intuitive discriminators of initiation signatures. We use $P(\cdot)$ to denote the power (as a function of time) and T to represent the second when a call is initiated. We divide a contiguous period of time around T into three periods: the first period of call initiation consisting of a window of w seconds including and after T , a window of w_- seconds prior to this period, and, a window of w_+ seconds after the call initiation period. We calculate the average power in each of these three periods and define the criteria for the three discriminators as follows:

1. The difference between the power in the call initiation period and the power in the period before call initiation is larger than a threshold τ_1 :

$$\overline{P([T, T + w - 1])} - \overline{P([T - w_-, T - 1])} \geq \tau_1 \quad (\text{B.1})$$

2. The difference between the power during the call initiation period and the power thereafter is larger than a threshold τ_2 :

$$\overline{P([T, T + w - 1])} - \overline{P([T + w, T + w + w_+ - 1])} \geq \tau_2 \quad (\text{B.2})$$

3. The difference between the power after and before call initiation is larger than a threshold τ_3 :

$$\overline{P([T + w, T + w + w_+ - 1])} - \overline{P([T - w_-, T - 1])} \geq \tau_3 \quad (\text{B.3})$$

Note that an advantage of using a window of several seconds in each period might be the potential reduction in the variance of the estimated power in that period. At

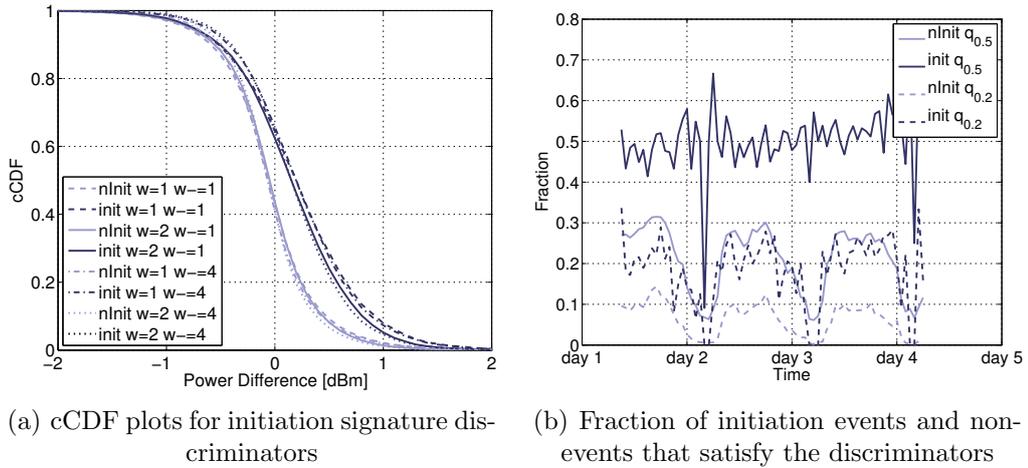


Figure B.10: Dataset $D1$: cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 1. The dark lines represent the distribution of the difference used by each criterion when a call is initiated and the light lines show the distribution when a call is not initiated.

the same time, larger periods may be polluted by other call initiation and termination events.

To better understand the impact of the various parameters including the window sizes and thresholds, we rely on so-called complementary Cumulative Distribution Function (cCDF) plots. Consider, e.g., the first discriminator. Recall that it looks at the difference during and before call initiation and expects this difference to exceed a threshold when a call is initiated. A cCDF plot shows if this discriminator is justified by plotting the distribution of the difference for all T when a call was initiated, and, compares it with the distribution for all T when a call was not initiated. Since we are interested in the number of cases that the difference exceeds a threshold, we plot the cCDF (defined as $1 - \text{CDF}$).

We show cCDF plots for several choices of w and w_- in Figure B.10(a). We see that if τ_1 is chosen to be 0.7 dBm and $w = w_- = 1$, then 20% of call initiation events satisfy the criterion of the first discriminator. This criterion is satisfied in less than 5% of the cases when no call was initiated. Thus, the probability that a call initiation satisfies the first criterion (with $\tau_1 = 0.7$ dBm) is about 4 times the probability compared to a randomly chosen second that sees no call initiation.

Figure B.10(a) also allows us to understand the choice of τ_1 . The higher we choose τ_1 to be, the fewer call initiation events satisfy the criterion. However, as we make τ_1 smaller, more seconds during which no call was initiated satisfy the criterion. The sweet spot appears to be at around 0.4 – 0.5 dBm, below which the latter increase faster than the former.

Figure B.10(a) also shows that we get marginally better results when we define the period during call initiation as consisting of exactly 1 second. Larger windows do not

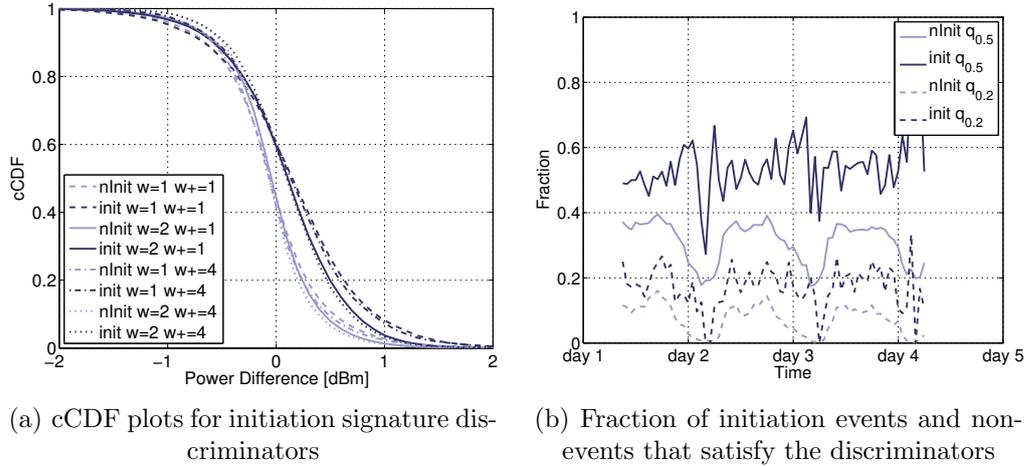


Figure B.11: Dataset $D1$: cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 2. The dark lines represent the distribution of the difference used by each criterion when a call is initiated and the light lines show the distribution when a call is not initiated.

improve our results. It seems that the power during additional seconds is not as high as that of the first second. Hence, any potential variance reduction from the additional power measurements comes at the cost of eliminating the signature itself. The size w_- of the period before call initiation impacts the results to a lesser extent, leading to marginally better results with a window of size 4.

In Figure B.11(a) and Figure B.12(a), we show the cCDF plots for the other two discriminators as well. We see that they, too, exhibit similar behavior, namely, smaller windows are better. $0.3 - 0.5$ dBm appears to be a good threshold value of τ_2 as well. However, the third criterion is not as useful since the difference in power before and after call initiation is not as clear.

The three plots show the potential behind using our three criteria as discriminators for call initiation events. However, their success may vary with time. We explore the time dependence in Figure B.10(b) for criterion 1, Figure B.11(b) for criterion 2, and Figure B.12(b) for criterion 3, respectively. We choose the thresholds to be the 0.5 and 0.2 cCDF quantiles from the corresponding cCDF figures. We then apply the criteria with the respective thresholds on an hourly basis and plot the fraction of initiation events and “non-events” satisfying them.

As expected, the fraction of initiation events satisfying each criterion does not vary significantly and stays around the quantile value (0.5 or 0.2) used to choose the thresholds. However, the fraction of “non-events” passing the criteria show a clear dip during night time for both thresholds. This implies significantly better performance during night time on account of lesser noise in power.

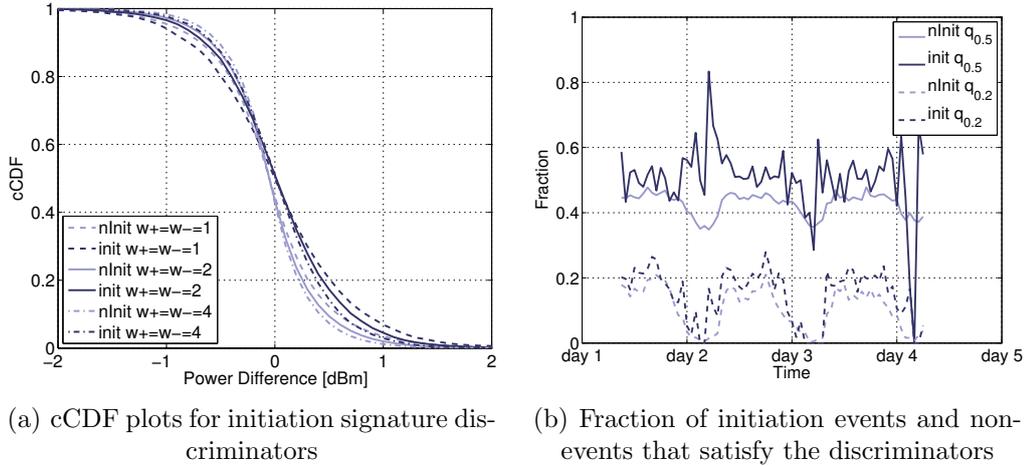


Figure B.12: Dataset $D1$: cCDF plots for initiation signature discriminators and fraction of initiation events and non-events that satisfy the discriminators for criterion 3. The dark lines represent the distribution of the difference used by each criterion when a call is initiated and the light lines show the distribution when a call is not initiated.

B.4.2 Initiation Detectors

We now investigate how the various discriminators and their criteria can be used for the best possible detection of initiation events. To quantify performance of such detection, we start by looking at the *detection probability*, i.e., the probability of detecting an initiation event given that a call was really initiated.

In Figure B.13(a), we plot the detection probability for each of our three discriminators used on dataset $D1$. For each detector, we vary the corresponding threshold from -1 to 2.5 dBm. As expected, the detection probability reduces with increasing thresholds. In Figure B.13(b), we show the probability of detecting a call initiation event, although no call was initiated for our three discriminators using $D1$.

Comparing Figure B.13(a) and Figure B.13(b) shows a common problem using energy detection. Choosing a threshold τ_1 that achieves a high detection probability results in many non-initiation events to be declared as initiation event and vice versa. E.g., using criterion 1, a threshold of $\tau_1 = 0$ dBm results in detecting close to 75% of the initiation events, but also in mistakenly detecting about 50% of the non-events as initiations. In contrast, a higher threshold of $\tau_1 = 0.5$ dBm detects only 10% of the non-events as initiations but fails to detect 70% of the initiations.

B.4.3 Discriminators of Termination Signature

Recall from Figure B.7 that the call termination events also have a signature, namely, the reduction of sensed power. These *t-average* plots motivate us to look at a natural criterion for a termination discriminator. This criterion would use two time periods: before and after call termination. The idea is to check if the power before any second T

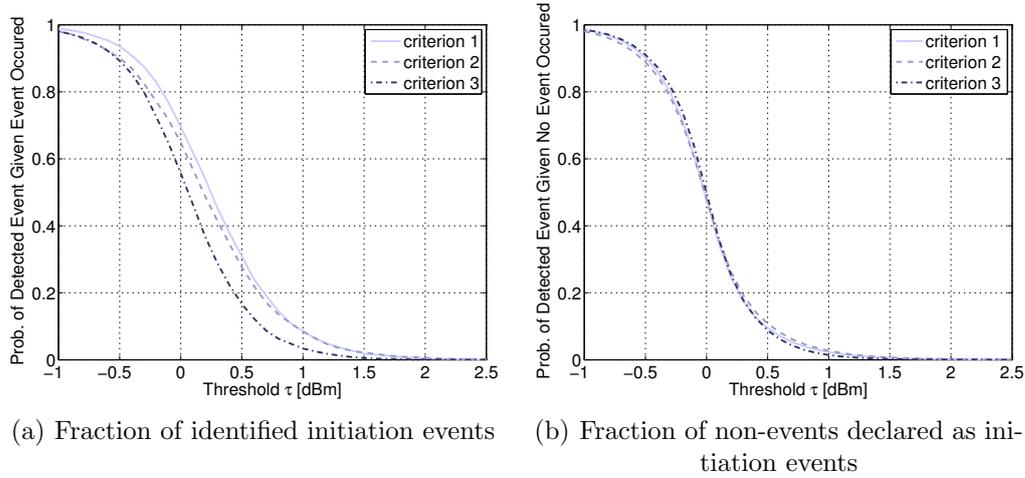


Figure B.13: Dataset $D1$: Fraction of identified initiation events and false positives by the detectors based on the 3 criteria with varying threshold values.

is greater than after. Formally, the criterion averages power in the two time periods of size w'_- and w'_+ and checks if:

$$\overline{P([T - w'_-, T - 1])} - \overline{P([T, T + w'_+ - 1])} \geq \tau' \quad (\text{B.4})$$

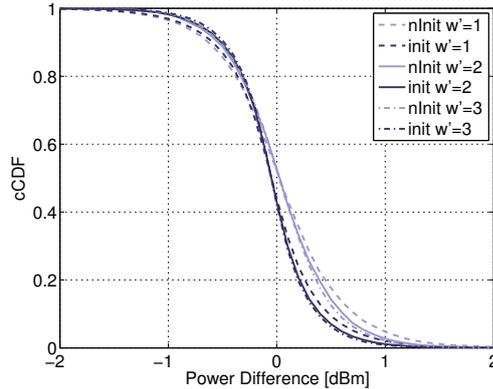
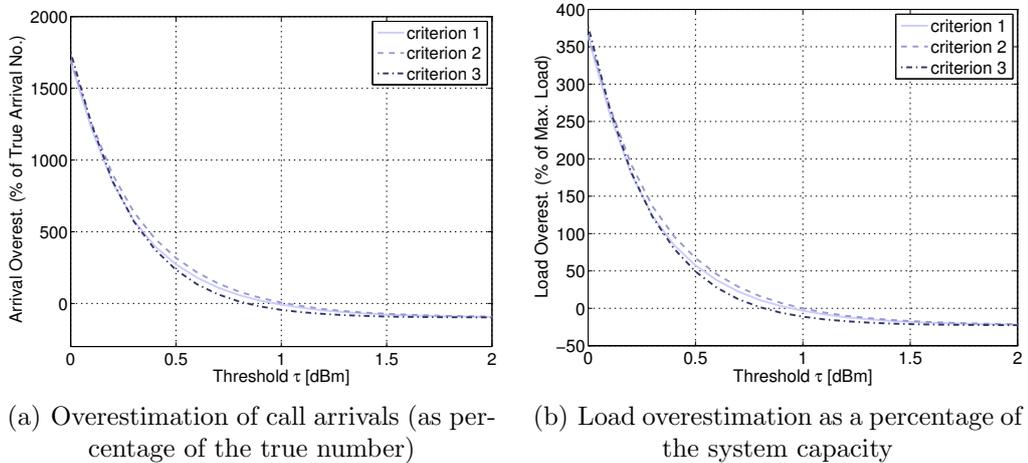
Note that, for consistency, we have phrased the criterion such that we continue to expect a higher value when an event occurs than when it does not.

In Figure B.14, we show the cCDF plot for this criterion. As before, we compare the cCDFs of the above difference for time instants during which calls terminated and the rest. We see that the criterion is not as clear-cut as for initiation events. In fact, the best τ' appears to be at around 0.6 dBm where termination events pass the criterion with twice the likelihood of other seconds. For the call termination criterion, we find that the smaller window size performs marginally better. All in all, we find that the discriminators of termination events do not perform in a promising way. Hence, we do not focus on event termination in the remainder of this investigation.

To summarize, both for initiation and termination events, though distinct signatures of power characteristics exist, we find that a single sensor encounters too much noise for accurate detection.

B.5 Estimating Unused Capacity

In this section, we discuss how we can use power sensed by a single sensor to derive useful estimates of unused capacity. Specifically, we find lower bounds for the unused capacity so that — at least — low-bandwidth applications can utilize it. Note that these estimates are useful because they are significantly greater than zero.


 Figure B.14: Dataset $D1$: cCDF plot for call termination.


(a) Overestimation of call arrivals (as percentage of the true number) (b) Load overestimation as a percentage of the system capacity

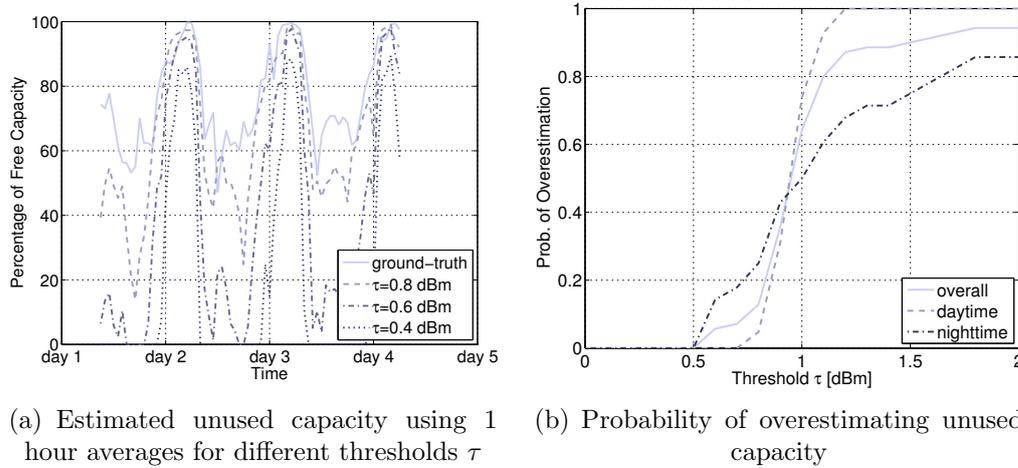
 Figure B.15: Dataset $D1$: Call arrival and load overestimation

As shown in the previous section, it is a lot harder to detect call termination events compared to detecting call initiation events. We, thus, only use our detectors to estimate call arrival rates. We then estimate the load in the system using Little's law as

$$E[k] = \lambda E[b] , \quad (\text{B.5})$$

where $E[b]$ is the mean call duration and λ the estimated arrival rate. We do assume that we have **partial** information about the system being studied, namely, the average call durations. Such information can be obtained from previous studies [64, 170] (since mean call durations are quite stable over time) or directly from providers. Since we use such information, we refer to it as a *gray-box* approach (in-between *black-box* and *white-box* approaches).

A crucial condition to not overestimate the unused capacity is to never *underestimate* the actual usage of the spectrum. In Figure B.15(a) we plot the difference between the


 Figure B.16: Dataset $D1$: (Over)estimation of unused capacity.

number of detected initiation events and the real initiation events using our three criteria and dataset $D1$. We plot this as a percentage of the number of real initiation events. Note that the detected initiation events can also include those that were mistakenly detected even though there was no underlying initiation event. Figure B.15(a) shows that we overestimate the primary usage whenever choosing threshold values $\tau < 1$ dBm (for all criteria). We observe similar results for $D2$ as well.

Using the (overestimated) number of call arrivals and mean call durations, we can estimate the load due to primary usage. We expect this to also be higher than the true load. We verify that this is true in Figure B.15(b) where we plot the amount by which the estimated load of 1-hour intervals exceeds the true load. Given our interest in unused capacity, we plot the load overestimation as a percentage of the maximum cell capacity. As with arrivals, with threshold values below 1 dBm, we overestimate load. For example, with 0.4 dBm, we overestimate load by 100% of the maximum load, on average.

If we calculate unused capacity by subtracting the (over)estimated load from the system capacity, we expect to *underestimate* the unused capacity. We show that this is indeed the case in Figure B.16(a). In this figure, we plot the true and estimated unused capacity for different threshold values using criterion 1. We again plot our results as a percentage of the system capacity. The figure shows that — in general — during the night the estimates are closer to the true available capacity than during the day. The figure also shows that with a threshold of 0.8 dBm we sometimes overestimate the unused capacity, whereas for a threshold of 0.4 dBm we often obtain estimates of zero capacity. An intermediate threshold of 0.6 dBm appears well-suited to be conservative and useful.

In Figure B.16(b) and Figure B.17 we finally show how well we can achieve our goal: finding unused spectrum to satisfy the secondaries bandwidth requirement with the constraint of not overestimating unused capacity. We show results for dataset $D1$ and criterion 1. We divide our dataset into hour-long time periods and calculate average

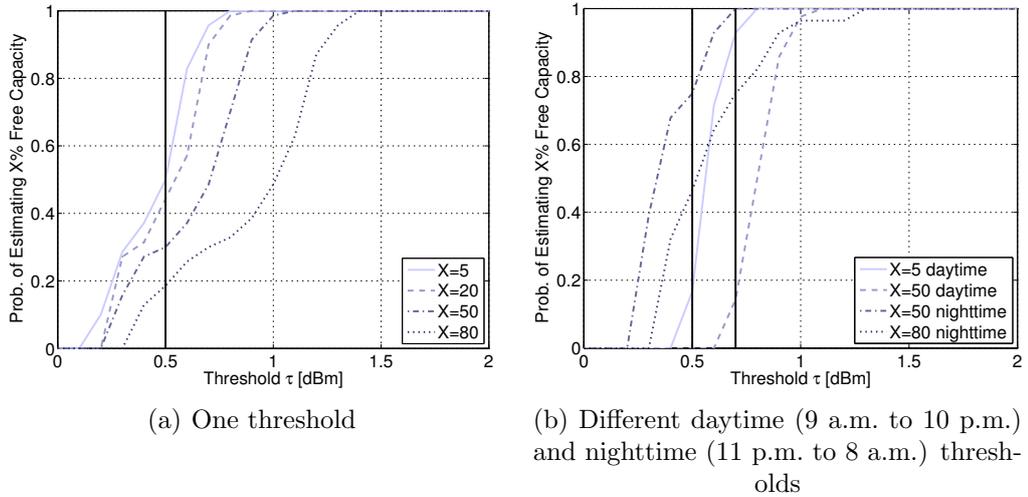


Figure B.17: Dataset $D1$: Probability of estimating at least X % of unused capacity.

values within these time periods.

Figure B.16(b) shows the probability that we overestimate the unused capacity using the whole dataset (solid line). For the SUs to not affect PUs, we need to choose thresholds where this line is zero. Consequently, we can only use thresholds $\tau \leq 0.5$ dBm. In Figure B.17(a), we show the probability of correctly estimating that at least X % of the capacity can be used by SUs. The horizontal solid black line in this figure indicates the maximum threshold (0.5 dBm) so that we *never* overestimate. Using this threshold, we are able to correctly find $X = 5$ % of capacity unused in more than half of the time periods.

If we use different thresholds for daytime (9 a.m. to 10 p.m.) time periods and nighttime (11 p.m. to 8 a.m.) time periods, we can improve performance significantly. As shown in Figure B.16(b), the maximum thresholds so that we never overestimate are $\tau \leq 0.7$ dBm for daytime and $\tau \leq 0.5$ dBm for nighttime, respectively. These are indicated by the horizontal solid black lines in Figure B.17(b). This figure also shows that, using the maximum daytime threshold ($\tau = 0.7$ dBm), we correctly estimate 5 % (50 %) of total capacity is unused in almost 100 % (14 %) of the daytime time periods. For the maximum nighttime threshold ($\tau = 0.5$ dBm) we correctly estimate that 50 % (80 %) of total capacity is unused in 75 % (46 %) of the nighttime hours.

For dataset $D2$, the results are similar though the maximum thresholds (so that we never overestimate) are closer to 1 dBm. These results show that power measurements at a single sensor can be quite useful especially for secondary applications such as urban sensing, which have low bandwidth demands but strict power constraints. Given the differences we observed between $D1$ and $D2$, local calibration might be necessary to choose the appropriate power thresholds.

B.6 Towards Spatial Diversity

In the previous section, we discussed how very noisy measurement data can nevertheless be used to derive meaningful lower bounds on unused capacity. An alternative approach is to improve the accuracy of event detection. Clearly, this is hard to achieve with a single sensor because the signatures of individual events are obfuscated by additive white noise, which makes accurate detection very hard especially in the low SNR regime [136, 147]. The natural way to improve event detection accuracy would be to use multiple sensors in spatially diverse locations so that we can eliminate the white noise. There are various proposals for cooperative spectrum sensing approaches in the literature, e.g., [111, 127, 144].

We now use our dataset to quantify the benefits of cooperative sensing. In particular, we consider the approach of soft decision combining as described in [111] and make a natural assumption of zero-mean additive Gaussian noise with variance σ^2 . With k distributed (and independent) sensors, the white noise in the average sensed power can be approximated as a zero-mean normal variable $\mathcal{N}(0, \frac{\sigma^2}{k})$ with variance $\frac{\sigma^2}{k}$ by the Central Limit Theorem [128].

We calculate the quality of event detection under the above model. Assume that a call is initiated at time T and we use the first criterion of Section B.4 with $w = w_- = 1$. At time T , the criterion computes the power spike of about 0.4 dBm plus the difference between the white noise at T and $T-1$. Assuming that white noise at these time instants is independent, their difference is a zero-mean normal variable with variance $\frac{2\sigma^2}{k}$. Hence, the criterion will have a false negative with probability P_{fn}

$$P_{\text{fn}} = P\left(\mathcal{N}\left(0, \frac{2\sigma^2}{k}\right) + 0.4 \leq \tau\right), \quad (\text{B.6})$$

where τ is the decision threshold. Conversely, if there is no call initiation at time t , the probability of a false positive P_{fp} is

$$P_{\text{fp}} = P\left(\mathcal{N}\left(0, \frac{2\sigma^2}{k}\right) \geq \tau\right). \quad (\text{B.7})$$

Given a maximum tolerable false positive and false negative probability, we can solve the above to estimate the minimum number of sensors (k) required. For example, if we desire no more than a fraction p of false positives *and* n of false negatives, then:

$$\begin{aligned} P_{\text{fn}} \leq p \quad \text{and} \quad P_{\text{fp}} \leq n \\ \implies \sqrt{k} \geq \max\left(\left|\frac{q_{1-p}\sqrt{2}\sigma}{\tau_1}\right|, \left|\frac{q_n\sqrt{2}\sigma}{\tau_1 - 0.4}\right|\right) \end{aligned} \quad (\text{B.8})$$

Here, q_y is the value at which the quantile of the standard normal distribution is y . We estimate σ using $D1$. Since we do not want to capture temporal mean variations, we remove the moving average of the previous 5 seconds for each power reading. This yields an estimate $\sigma = 0.39$ with which we can achieve less than 10% false negatives and false

positives using $k = 13$ sensors and $\tau = 0.2$ dBm. In the previous section, we saw that it may be desirable to have a much smaller fraction of false negatives than false positives. It turns out that we can achieve at most 1 % false negatives and 10 % false positives with $k = 26$ sensors and $\tau = 0.25$ dBm. Deploying such numbers of sensors per cell could very well be economical, especially if existing consumer devices can be leveraged. Of course, multiple sensors are likely to yield more accurate results than the lower bounds of the previous section.

B.7 Related Work

In [106], general guidelines are given on how to perform spectrum measurements. The authors performed a measurement campaign in the US from 1992 [162] where they actually found the usage of the ISM band at 2.4 GHz very low. Fleurke et al. [51] differentiate between the actual occupancy and the mean occupancy. They describe different sampling methods (random and systematic sampling). In [133] a detailed modeling proposal for WLAN and microwave ovens is given.

Measuring and modeling the spectrum occupancy in the high frequency (HF) range (3 – 30 MHz) had drawn a lot of research attention since the late 1970's. Various models of the boolean variable that represents whether or not a channel is used have been proposed. Most of these are tailored to the specific characteristics and usage patterns in the HF spectrum. The Laycock-Gott logit model [44, 61, 94, 119] requires 25 parameters to model the measured data. Measurements are taken during the solstices, where atmospheric and galactic noise is low. Goultelard et al. [8] have developed a model for estimating the time variations of the so called congestion (occupancy) using channel availability statistics. In [140], the authors develop guidelines on sensing times and estimate confidence limits for these times. They also propose using first-order Markov chains to model channel availability. This work has been extended in [59, 60] to two-dimensional first-order Markov chains and in [124] to cyclostationary two-dimensional first-order Markov chains.

In recent years, a lot of measurement studies [27, 102, 109] have been carried out to show the under-utilization of the licensed spectrum. Though these studies show the abundance of temporally unused spectrum, they give little insight into the dynamic behavior of the licensed users legally operating in those bands.

A measurement campaign focusing on the cellular voice bands was carried out during the soccer world-cup 2006 in Germany [69, 132]. The authors show the differences in spectrum occupancy in the GSM and UMTS bands before, during, and after a match. the authors correlate differences in the power measurements to beginning, half time, and end of the match. However, no detailed load analysis is performed. In contrast, the authors of [89] estimate the load in the New York cellular bands (CDMA as well as GSM) based on spectrum measurements. However, in addition to pure power measurements, the CDMA signals are demodulated to determine the number of active Walsh codes (i.e., the number of ongoing calls). To determine the number of calls in the GSM bands, image processing of the spectrogram snapshots is used. This is in contrast to our study, which

is based on power measurements and uses minimal processing. In addition, to our best knowledge, there is no study which correlates spectrum measurements with the actual load as recorded by the system. Sensing for TV bands has been previously studied, for example in [91]. They found that energy detection with multiple sensors is often better than feature detection. However, their results are for the relatively static TV bands and not for cellular bands.

B.8 Conclusions

We used a unique set of simultaneous sensing and network measurements to study the problem of sensing-based estimation of unused capacity in cellular spectrum. When averaged over time, we found well-defined signatures of call initiation and termination events using the power at a single sensor. However, sensing noise makes it challenging to use these signatures to estimate unused capacity by identifying call events. We found that useful underestimates can nevertheless be computed especially for low-bandwidth secondary applications. Alternatively, we can obtain accurate estimates by using multiple sensors. To our knowledge, our work is the first detailed study of how well sensing works in CDMA networks, which are often viewed as candidates for DSA. In the future, we intend to design and evaluate better sensing algorithms including those based on multiple sensors.

APPENDIX

Acronyms

AP Access Point

ATSC Advanced Television System Committee

cCDF complementary Cumulative Distribution Function

CDF Cumulative Distribution Function

CDMA Code Division Multiple Access

CEPT European Conference of Postal and Telecommunications Administrations

CORVUS Cognitive Radio for usage of Virtual Unlicensed Spectrum

CPE Customer Premise Equipment

CR Cognitive Radio

CSMA Carrier Sense Multiple Access

DARPA U.S. Defense Advanced Research Projects Agency
(www.darpa.mil)

DECT Digital Enhanced Cordless Telecommunications

DFH Dynamic Frequency Hopping

DSA Dynamic Spectrum Access

EHF Extremely High Frequency

EU European Union
(www.europa.eu)

FCC U.S. Federal Communications Commission
(www.fcc.gov)

FEC Forward Error Correction

FM Frequency Modulation

HF High Frequency

IEEE Institute of Electrical and Electronics Engineers
(www.ieee.org)

i.i.d. independent and identically distributed

IS-95 Interim Standard 95

ISM Industrial Scientific and Medical

ITU-R International Telecommunication Union Radiocommunication Sector
(www.itu.int/ITU_R/)

GSM Global System for Mobile Communications

LF Low Frequency

LOS line-of-sight

LTE Long Term Evolution

MAC Medium Access Control

MF Medium Frequency

MIMO Multiple Input Multiple Output

NLOS non-line-of-sight

NTSC National Television System Committee

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

PDF Probability Density Function

PSD Power Spectral Density

PSK Phase-Shift Keying

PU Primary User

PU band spectrum band owned by a Primary User

QoS Quality of Service

RAN Radio Access Network

RF Radio Frequency

ROC Receiver Operating Characteristic

RSSI Received Signal Strength Indication

SCC41 IEEE Standards Coordinating Committee 41

SDR Software Defined Radio

SHF Super High Frequency

SNR Signal-to-Noise Ratio

SU Secondary User

SUG Secondary User Group

SUL Secondary User Link

TDMA Time Division Multiple Access

UHF Ultra High Frequency

UMTS Universal Mobile Telecommunications System

UNII Unlicensed National Information Infrastructure

USB Universal Serial Bus

UWB Ultra-Wideband

VHF Very High Frequency

VLf Very Low Frequency

WiMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

WRAN Wireless Regional Area Network

APPENDIX **D**

Variables and Symbols

B CR system bandwidth

b_{pu} PU bandwidth

b_{sc} OFDM sub-carrier bandwidth

b_{sub} sub-channel bandwidth

b_{sul} SUL bandwidth

b_{sens} sensing bandwidth

b_{coh} coherence bandwidth

N_{pu} number of PU bands in the CR system

N_{ofdm} number of OFDM sub-carriers of the CR system

N_{sub} number of sub-channels of a SUL

N_{sc} number of OFDM sub-carriers per sub-channel

N_{s} number of micro-samples used for energy detection

X number of redundant / reserved sub-channels of a SUL

X_{opt} optimum number of redundant / reserved sub-channels of a SUL

t_{max} maximum interference time

t_{sens}	local sensing time
t_{coop}	distributed sensing time
t_{reconf}	link reconfiguration time
t_{coh}	coherence time
t_{slot}	slot length
c	speed of light
f_c	carrier frequency
f_d	maximum Doppler shift
v	average velocity
σ_{ds}	standard deviation of the delay spread
f_s	sampling frequency
$\rho_{\mathbf{x},\mathbf{y}}$	Person product-moment correlation coefficient
P_{fp}	probability of false positives
P_{fn}	probability of false negatives
$\overline{P_{\text{fp}}}$	probability of false positives of the SUG
$\overline{P_{\text{fn}}}$	probability of false negatives of the SUG
P_{detect}	probability that a PU is detected within its PU band
P_{qos}	probability of violating the QoS requirement of the SUL
μ_{pu}	usage duration
λ_{pu}	arrival rate
L_{pu}	PU system load
L_{sens}	sensed PU system load
P_{pu}	probability of PU presence
$\Phi_{i,t}$	Power Spectral Density of spectrum band i at time t
$\Psi_{i,t}$	estimated binary spectrum occupancy variable of spectrum band i at time t
τ	detection threshold
D_i	estimated average duty cycle of spectrum band i

$\sigma_{\mathbf{n}}^2$ noise variance
 $\sigma_{\mathbf{s}}^2$ signal variance
 h_k channel coefficient of k^{th} mini-block
 M number of sensing samples
 M_t number of sensing samples in the time domain
 M_b number of sensing samples in the frequency domain
 m number of cooperating sensors
 γ received SNR
 d_d diversity difference fraction
 C link capacity
 C_{avg} average link capacity
 C_{max} maximum link capacity
 C_α α quantile of the link capacity
 C_i link capacity in slot i
 S_{max} theoretically available spectrum
 S_{sens} sensed available spectrum
 $S_{\text{eff}}^{\text{sens}}$ effectively available spectrum after spectrum sensing
 S_{eff} effectively available spectrum

APPENDIX E

Publications

Below is a list of publications of the author generated within the time the presented thesis was done. Within each paragraph, the publications are sorted in chronological order.

Journal Papers:

D. Willkomm and A. Wolisz. Efficient QoS support for secondary users in cognitive radio systems. *IEEE Wireless Communications Magazine*, 17(4):16–23, Aug. 2010

D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary user behavior in cellular networks and implications for dynamic spectrum access. *IEEE Communications Magazine*, 47(3):88–95, Mar. 2009

W. Hu, D. Willkomm, L. Chu, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz. Dynamic frequency hopping communities for efficient IEEE 802.22 operation. *IEEE Communications Magazine*, 45(5):80–87, May 2007

Book Chapter:

A. de Baynast, M. Bohge, D. Willkomm, and J. Gross. *Modeling and Tools for Network Simulation*, chapter Physical Layer Modeling, pages 135–172. Springer, 1 edition, Mar. 2010

Conference Proceedings:

- D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. The problem of sensing unused cellular spectrum. In *Proc. of the 10th International IFIP TC 6 Networking Conference*, Lecture Notes in Computer Science, May 2011
- C. Dombrowski, D. Willkomm, and A. Wolisz. Is high quality sensing really necessary for opportunistic spectrum usage? In *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '10)*, Dec. 2010
- D. Willkomm, W. Hu, D. Hollos, J. Gross, and A. Wolisz. On centralized and distributed frequency assignment in cognitive radio based frequency hopping cellular networks. In *In Proc. of the 3rd International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART '10)*, Nov. 2010. Invited paper
- D. Willkomm and A. Wolisz. Is oversensitive spectrum sensing the door opener for initial cognitive radio deployments? In *Proc. of the 2nd ACM MobiHoc S3 Workshop*, Sept. 2010
- L. M. Feeney and D. Willkomm. Energy framework: An extensible framework for simulating battery consumption in wireless networks. In *Proceeding of the 3. International Workshop on OMNeT++*, pages 1–4, Mar. 2010
- K. Wessel, M. Swigulski, A. Köpke, and D. Willkomm. MiXiM - the physical layer: An architecture overview. In *Proceeding of the 2. International Workshop on OMNeT++*, pages 1–8, Mar. 2009
- D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary users in cellular networks: A large-scale measurement study. In *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '08)*, Chicago, IL, Oct. 2008. Winner of best paper award
- D. Willkomm, M. Bohge, D. Hollós, J. Gross, and A. Wolisz. Double hopping: A new approach for dynamic frequency hopping in cognitive radio networks. In *Proc. of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, Sept. 2008
- A. Köpke, M. Swigulski, K. Wessel, D. Willkomm, P. T. Klein Haneveld, T. E. V. Parker, O. W. Visser, H. S. Lichte, and S. Valentin. Simulating wireless and mobile networks in OMNeT++ the MiXiM vision. In *Proceeding of the 1. International Workshop on OMNeT++*, Mar. 2008
- D. Willkomm, J. Gross, and A. Wolisz. Reliable link maintenance in cognitive radio systems. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 371–378, Baltimore (MD), Nov. 2005

S. M. Mishra, D. Cabric, C. Chang, D. Willkomm, B. van Schewick, A. Wolisz, and R. W. Brodersen. A real time cognitive radio testbed for physical and link layer experiments. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 562–567, Baltimore (MD), Nov. 2005

D. Cabric, S. M. Mishra, D. Willkomm, R. W. Brodersen, and A. Wolisz. A cognitive radio approach for usage of virtual unlicensed spectrum. In *Proc. of the 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005

Standardization:

L. Chu, W. Hu, G. Vlantis, J. Gross, M. Abusubaih, D. Willkomm, and A. Wolisz. Dynamic frequency hopping community. Technical proposal submitted to IEEE 802.22 WG 22-06-0113, IEEE 802.22 Working Group, June 2006

TKN Research Report Series

D. Willkomm. Spectrum efficient QoS support for secondary users in cognitive radio systems. Technical Report TKN-09-008, Telecommunication Networks Group, Technische Universität Berlin, Dec. 2009

D. Willkomm, M. Bohge, D. Hollos, and J. Gross. Double hopping: A new approach for dynamic frequency hopping in cognitive radio networks. Technical Report TKN-08-001, Telecommunication Networks Group, Technische Universität Berlin, Jan. 2008

D. Hollos, D. Willkomm, J. Gross, and W. Hu. Centralized vs. distributed frequency assignment in frequency hopping (cognitive radio) cellular networks. Technical Report TKN-07-007, Telecommunication Networks Group, Technische Universität Berlin, Dec. 2007

Miscellaneous:

R. W. Brodersen, A. Wolisz, D. Cabric, S. M. Mishra, and D. Willkomm. CORVUS: A cognitive radio approach for usage of virtual unlicensed spectrum. White Paper, 2004

Bibliography

- [1] 3GPP. Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-advanced). Technical Report TR 36.913, Technical Specification Group Radio Access Network, 2009.
- [2] Agilent. W1314a datasheet. www.agilent.com, 2008. [online].
- [3] A. Aguiar and J. Gross. Wireless channel models. Technical Report TKN-03-007, Telecommunication Networks Group, Technische Universität Berlin, apr 2003.
- [4] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks*, 50(13): 2127–2159, September 2006.
- [5] A. Al Daoud, M. Alanyali, and D. Starobinski. Secondary pricing of spectrum in cellular CDMA networks. In *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pages 535–542, 2007.
- [6] G. Alyfantis, G. Marias, S. Hadjiefthymiades, and L. Merakos. Non-cooperative dynamic spectrum access for cdma networks. In *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '07)*, pages 3574–3578, 26–30 Nov. 2007.
- [7] Analysys, DotEcon, and Hogan & Hartson. Study on conditions and options in introducing secondary trading of radio spectrum in the european community, May 2004.
- [8] C. and Goutelard and J. Caratori. Time modelling of H.F. interferences. In J. Caratori, editor, *Proceedings of the Fifth International Conference on HF Radio Systems and Techniques*, pages 343–348, 1991.
- [9] D. D. Ariananda, M. K. Lakshmanan, and H. Nikoo. A survey on spectrum sensing techniques for cognitive radio. In *Proc. of the 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART '09)*, pages 74–79, May 18–20, 2009.
- [10] S. Atapattu, C. Tellambura, and H. Jiang. Performance of an energy detector over channels with both multipath fading and shadowing. *IEEE Transactions on Wireless Communications*, 9(12):3662–3670, 2010.
- [11] R. Bacchus, A. Fertner, C. Hood, and D. Roberson. Long-term, wide-band spectral monitoring in support of dynamic spectrum access networks at the iit spectrum observatory. In *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '08)*, pages 1–10, Oct. 2008.
- [12] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh. White space networking with wi-fi like connectivity. In *Proceedings of the ACM SIGCOMM conference on Data communication (SIGCOMM 2009)*, pages 27–38, New York, NY, USA, 2009. ACM.
- [13] L. Berlemann, S. Mangold, and B. H. Walke. Policy-based reasoning for spectrum sharing

- in radio networks. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 1–10, Nov. 2005.
- [14] V. Blaschke, H. Jaekel, T. Renk, C. Kloeck, and F. Jondral. Occupation measurements supporting dynamic spectrum allocation for cognitive radio design. In *Proc. of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '07)*, Aug. 2007.
- [15] M. Bohge. *Dynamic Resource Allocation in Packet-Oriented Multi-Cell OFDMA Systems*. PhD thesis, Technische Universität Berlin, 2010.
- [16] M. Bohge, J. Gross, M. Meyer, and A. Wolisz. Dynamic resource allocation in OFDM systems: An overview of cross-layer optimization principles and techniques. *IEEE Network*, 21(1):53–59, Jan. 2007.
- [17] R. W. Broderon, A. Wolisz, D. Cabric, S. M. Mishra, and D. Willkomm. CORVUS: A cognitive radio approach for usage of virtual unlicensed spectrum. White Paper, 2004.
- [18] M. M. Buddhikot and K. Ryan. Spectrum management in coordinated dynamic spectrum access based cellular networks. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 299–307, Baltimore (MD), Nov. 2005.
- [19] M. M. Buddhikot, P. Kolodzy, S. Miller, K. Ryan, and J. Evans. DIMSUMNet: New directions in wireless networking using coordinated dynamic spectrum access. In *Proceedings of the Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM 05)*, pages 78–85, June 2005.
- [20] D. Cabric, S. M. Mishra, D. Willkomm, R. W. Broderon, and A. Wolisz. A cognitive radio approach for usage of virtual unlicensed spectrum. In *Proc. of the 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.
- [21] D. Cabric, A. Tkachenko, and R. W. Brodersen. Spectrum sensing measurements of pilot, energy, and collaborative detection. In *Proc. of the IEEE Military Communications Conference (MILCOM '06)*, pages 1–7, 2006.
- [22] F. Capar and F. Jondral. Resource allocation in a spectrum pooling system for packet radio networks using OFDM/TDMA. In *Proc. of the IST Mobile & Wireless Telecommunications Summit*, pages 489–493, Thessaloniki, Greece, 2002.
- [23] J. Cavers. *Mobile Channel Characteristics*. Kluwer Academic Publishers, 2000.
- [24] D. Chen, S. Yin, Q. Zhang, M. Liu, and S. Li. Mining spectrum usage data: a large-scale spectrum measurement study. In *Proc. of the 15th ACM International Conference on Mobile Computing and Networking (MobiCom '09)*, pages 13–24, 2009.
- [25] H.-S. Chen and W. Gao. Spectrum sensing for tv white space in north america. *IEEE Journal on Selected Areas in Communications*, 29(2):316–326, 2011.
- [26] R. Chen, J.-M. Park, and K. Bian. Robust distributed spectrum sensing in cognitive radio networks. In *Proc. of the 27th IEEE International Conference on Computer Communications (INFOCOM '08)*, pages 1876–1884, 2008.
- [27] R. Chiang, G. Rowe, and K. Sowerby. A quantitative analysis of spectral occupancy measurements for cognitive radio. In *Proc. of the 65th IEEE Vehicular Technology Conference (VTC-Spring '07)*, pages 3016–3020, 2007.
- [28] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In *Proc. of the 16th ACM International Conference on Mobile Computing and Networking (MobiCom '10)*, pages 1–12, 2010.
- [29] Y.-J. Choi, Y. Xin, and S. Rangarajan. Overhead-throughput tradeoff in cooperative

- cognitive radio networks. In *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC '09)*, pages 1–6, Apr. 2009.
- [30] C.-T. Chou, S. S. N, H. Kim, and K. G. Shi. What and how much to gain by spectrum agility? *IEEE Journal on Selected Areas in Communications*, 25(3):576–588, apr 2007.
- [31] L. Chu, W. Hu, G. Vlantis, J. Gross, M. Abusubaih, D. Willkomm, and A. Wolisz. Dynamic frequency hopping community. Technical proposal submitted to IEEE 802.22 WG 22-06-0113, IEEE 802.22 Working Group, June 2006.
- [32] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar. IEEE 802.22: The first worldwide wireless standard based on cognitive radios. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 328–337, Baltimore (MD), Nov. 2005.
- [33] D. Danev. On signal detection techniques for the dvb-t standard. In *Proc. of the 4th International Symposium on Communications, Control and Signal Processing (ISCCSP '10)*, pages 1–5, 2010.
- [34] Darpa. Darpa neXt Generation program, 2003.
- [35] Darpa. The XG architectural framework RFC v1.0, 2003.
- [36] Darpa. The XG vision RFC v1.0, 2003.
- [37] A. de Baynast, M. Bohge, D. Willkomm, and J. Gross. *Modeling and Tools for Network Simulation*, chapter Physical Layer Modeling, pages 135–172. Springer, 1 edition, Mar. 2010.
- [38] R. de Lacerda, L. S. Cardoso, R. Knopp, D. Gesbert, and M. Debbah. EMOS platform: Real-time capacity estimation of MIMO channels in the UMTS-TDD band. In *Proc. of the 4th Int. Symp. Wireless Communication Systems (ISWCS 07)*, pages 782–786, 2007.
- [39] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank. Great expectations: The value of spatial diversity in wireless networks. *Proceedings of the IEEE*, 92(2):219–270, 2004.
- [40] F. F. Digham, M.-S. Alouini, and M. K. Simon. On the energy detection of unknown signals over fading channels. In *Proc. of the IEEE International Conference on Communications (ICC '03)*, volume 5, pages 3575–3579, 2003.
- [41] F. F. Digham, M.-S. Alouini, and M. K. Simon. On the energy detection of unknown signals over fading channels. *IEEE Transactions on Communications*, 55(1):21–24, 2007.
- [42] C. Dombrowski, D. Willkomm, and A. Wolisz. Is high quality sensing really necessary for opportunistic spectrum usage? In *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '10)*, Dec. 2010.
- [43] J. Eberspächer, H.-J. Vogel, C. Bettstetter, and C. Hartmann. *GSM, Architecture, Protocols and Services*. John Wiley, Ltd, 3 edition, 2001.
- [44] L. Economou, H. Haralambous, C. Pantjairois, P. Green, G. Gott, P. Laycock, M. Broms, and S. Boberg. Models of HF spectral occupancy over a sunspot cycle. *IEE Proceedings-Communications*, 152(6):980–988, 9 Dec. 2005.
- [45] Ettus Research LLC. Universal software radio peripheral product (usrp), 2011. URL <http://www.ettus.com/>.
- [46] FCC. ET docket no. 03-322. Notice of Proposed Rule Making and Order, December 2003.
- [47] FCC. ET docket no. 04-113. Notice of Proposed Rulemaking, May 2004.
- [48] FCC. Initial evaluation of the performance of prototype TV- band white space devices. OET Report FCC/OET 07-TR-1006, July 2007.
- [49] FCC. ET docket no. 10-174. Second Memorandum Opinion and Order, Sept. 2010.

- [50] L. M. Feeney and D. Willkomm. Energy framework: An extensible framework for simulating battery consumption in wireless networks. In *Proceeding of the 3. International Workshop on OMNeT++*, pages 1–4, Mar. 2010.
- [51] S. R. Fleurke, H. G. Dehling, H. K. Leonhard, A. D. Brinkerink, and R. den Besten. Measurement and statistical analysis of spectrum occupancy. *European Transactions on Telecommunications*, 15(5):429–436, 2004.
- [52] V. Gaddam and M. Ghosh. Robust sensing of dvt-t signals. In *Proc. of the 4th IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10)*, pages 1–8, 2010.
- [53] S. Geirhofer, L. Tong, and B. Sadler. Dynamic spectrum access in the time domain: Modeling and exploiting white space. *IEEE Communications Magazine*, 45(5):66–72, May 2007.
- [54] X. Gelabert, I. F. Akyildiz, O. Sallent, and R. Agustí. Operating point selection for primary and secondary users in cognitive radio networks. *Computer Networks*, 53(8):1158–1170, 2009.
- [55] GENI. GENI project, 2011. URL <http://www.geni.net>.
- [56] R. Gentleman and R. Ihaka. R project for statistical computing, v. 2.9.1, 2009. URL <http://www.r-project.org>.
- [57] A. Ghasemi and E. S. Sousa. Collaborative spectrum sensing for opportunistic access in fading environments. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 131 – 136, Baltimore (MD), Nov. 2005.
- [58] C. Ghosh, S. Pagadarai, D. Agrawal, and A. Wyglinski. A framework for statistical wireless spectrum occupancy modeling. *IEEE Transactions on Wireless Communications*, 9(1):38–44, Jan. 2010.
- [59] A. J. Gibson and L. Arnett. Statistical modelling of spectrum occupancy. *Electronics Letters*, 29(25):2175–2176, 1993.
- [60] A. J. Gibson and L. Arnett. Measurements and statistical modelling of spectrum occupancy. In *Proceedings of the 6th International Conference on HF Radio Systems and Techniques*, pages 150–154, 1994.
- [61] G. Gott, S. Chan, C. Pantjjaros, and P. Laycock. High frequency spectral occupancy at the solstices. *IEE Proceedings-Communications*, 144(1):24–32, Feb. 1997.
- [62] D. Grandblaise, D. Bourse, K. Moessner, and P. Leaves. Dynamic spectrum allocation (DSA) and reconfigurability. In *Proceedings of the Software-Defined Radio (SDR) Forum*, Nov. 2002.
- [63] D. Greenwood and L. Hanzo. *Mobile Radio Communications*, chapter Characterization of Mobile Radio Channels, pages 92–185. Pentech Press, 1992.
- [64] J. Guo, F. Liu, and Z. Zhu. Estimate the call duration distribution parameters in GSM system based on k-l divergence method. In *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007)*, pages 2988–2991, 2007.
- [65] B. Hamdaoui. Adaptive spectrum assessment for opportunistic access in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 8(2):922–930, Feb. 2009.
- [66] S. Hamouda and B. Hamdaoui. Dynamic spectrum access in heterogeneous networks: Hsdpa and wimax. In *IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing*, pages 1253–1257, New York, NY, USA,

2009. ACM.
- [67] Y. Han, Y. Wen, W. Tang, and S. Li. Spectrum occupancy measurement: Focus on the tv frequency. In *Proc. of the 2nd International Conference on Signal Processing Systems (ICSPS '10)*, volume 2, pages V2-490-V2-494, July 2010.
 - [68] S. P. Herath and N. Rajatheva. Analysis of equal gain combining in energy detection for cognitive radio over nakagami channels. In *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '09)*, pages 1-5, 2008.
 - [69] O. Holland, P. Cordier, M. Muck, L. Mazet, C. Klock, and T. Renk. Spectrum power measurements in 2G and 3G cellular phone bands during the 2006 football world cup in germany. In *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pages 575-578, Apr. 2007.
 - [70] D. Hollos, D. Willkomm, J. Gross, and W. Hu. Centralized vs. distributed frequency assignment in frequency hopping (cognitive radio) cellular networks. Technical Report TKN-07-007, Telecommunication Networks Group, Technische Universität Berlin, Dec. 2007.
 - [71] W. Hu, D. Willkomm, L. Chu, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz. Dynamic frequency hopping communities for efficient IEEE 802.22 operation. *IEEE Communications Magazine*, 45(5):80-87, May 2007.
 - [72] S. Huang, X. Liu, and Z. Ding. Optimal sensing-transmission structure for dynamic spectrum access. In *Proc. of the 28th IEEE International Conference on Computer Communications (INFOCOM '09)*, pages 2295-2303, Apr. 19-25, 2009.
 - [73] IEEE. IEEE 802.22 working group, 2010. URL <http://www.ieee802.org/22>.
 - [74] IEEE Std 1900.1-2008. Definitions and concepts for dynamic spectrum access: Terminology relating to emerging wireless networks, system functionality, and spectrum management. Technical report, IEEE, 2008.
 - [75] IEEE Std 1900.2-2008. Recommended practice for the analysis of in-band and adjacent band interference and coexistence between radio systems. Technical report, IEEE, 29 2008.
 - [76] IEEE Std 1900.4-2009. Architectural building blocks enabling network-device distributed decision making for optimized radio resource usage in heterogeneous wireless access networks. Technical report, IEEE, Feb. 2009.
 - [77] IEEE Std 802-11y-2008. Specific requirements: 3650-3700 mhz operation in usa. Technical report, IEEE, 2008.
 - [78] IEEE Std 802.11-2007. Part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. Technical report, IEEE, 2007.
 - [79] IEEE Std 802.11k-2008. Amendment 1: Radio resource measurement of wireless lans. Technical report, IEEE, 2008.
 - [80] IEEE Std 802.16-2009. Part 16: Air interface for broadband wireless access systems. Technical report, IEEE, 2009.
 - [81] IEEE Std 802.22.1-2010. Part 22.1: Standard to enhance harmful interference protection for low-power licensed devices operating in tv broadcast bands. Technical report, IEEE, 2010.
 - [82] IEEE Std P802.22/D6.1. Draft standard for wireless regional area networks part 22: Cognitive wireless ran medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands. Draft standard, IEEE, 2010.
 - [83] IEEE802.11b. Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 2.4 GHz band. Technical report,

- IEEE Computer Society, 2000.
- [84] O. Ileri, D. Samardzija, and N. Mandayam. Demand responsive pricing and competitive spectrum allocation via a spectrum server. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 194–202, 2005.
 - [85] M. Islam, C. Koh, S. Oh, X. Qing, Y. Lai, C. Wang, Y.-C. Liang, B. Toh, F. Chin, G. Tan, and W. Toh. Spectrum survey in singapore: Occupancy measurements and analyses. In *Proc. of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '08)*, pages 1–7, May 2008.
 - [86] P. Ji, Z. Ge, J. Kurose, and D. Towsley. A comparison of hard-state and soft-state signaling protocols. *IEEE/ACM Transactions on Networking*, 15(2):281–294, April 2007.
 - [87] F. Kaltenberger, L. Bernado, and T. Zemen. Characterization of measured multi-user MIMO channels using the spectral divergence measure. In *In Proc. of the COST 2100 6th Management Committee Meeting*, 2008.
 - [88] F. Kaltenberger, D. Gesbert, R. Knopp, and M. Kountouris. Correlation and capacity of measured multi-user MIMO channels. In *Proc. of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, pages 1–5, 2008.
 - [89] T. Kamakaris, M. M. Buddhikot, and R. Iyer. A case for coordinated dynamic spectrum access in cellular networks. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 289–298, Baltimore (MD), Nov. 2005.
 - [90] H. Kim and K. G. Shin. In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? In *Proc. of the 14th ACM International Conference on Mobile Computing and Networking (MobiCom '08)*, pages 14–25, New York, NY, USA, 2008. ACM.
 - [91] H. Kim and K. G. Shin. In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? In *Proc. of the 14th ACM International Conference on Mobile Computing and Networking (MobiCom '08)*, pages 14–25, 2008.
 - [92] A. Köpke, M. Swigulski, K. Wessel, D. Willkomm, P. T. Klein Haneveld, T. E. V. Parker, O. W. Visser, H. S. Lichte, and S. Valentin. Simulating wireless and mobile networks in OMNeT++ the MiXiM vision. In *Proceeding of the 1. International Workshop on OMNeT++*, Mar. 2008.
 - [93] J. Laiho, A. Wacker, and T. Novosad, editors. *Radio Network Planning and Optimisation for UMTS*. John Wiley & Sons, Inc., 2002.
 - [94] P. Laycock, M. Morrell, G. Gott, and A. Ray. A model for HF spectral occupancy. In *Proc. Fourth International Conference on HF Radio Systems and Techniques*, pages 165–171, 11–14 Apr 1988.
 - [95] P. Leaves, J. Huschke, and R. Tafazolli. A summary of dynamic spectrum allocation results from DRiVE. In *Proc. of the IST Mobile & Wireless Telecommunications Summit*, pages 245–250, 2002.
 - [96] P. Leaves, K. Moessner, R. Tafazolli, D. Grandblaise, D. Bourse, R. Tonjes, and M. Breveglieri. Dynamic spectrum allocation in composite reconfigurable wireless networks. *IEEE Communications Magazine*, 42:72–81, may 2004.
 - [97] W.-Y. Lee and I. F. Akyildiz. Optimal spectrum sensing framework for cognitive radio networks. *IEEE Transactions on Wireless Communications*, 7(10):3845–3857, Oct. 2008.
 - [98] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang. Sensing-throughput tradeoff for cognitive

- radio networks. In *Proc. of the IEEE International Conference on Communications (ICC '07)*, pages 5330–5335, 24–28 June 2007.
- [99] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang. Sensing-throughput tradeoff for cognitive radio networks. *IEEE Transactions on Wireless Communications*, 7(4):1326–1337, 2008.
- [100] J. D. C. Little. A proof for the queuing formula: $L = \lambda W$. *Operations Research*, 9(3): 383–387, 1961. URL <http://www.jstor.org/stable/167570>.
- [101] M. Lopez-Benitez, A. Umberto, and F. Casadevall. Evaluation of spectrum occupancy in Spain for cognitive radio applications. In *Proc. of the 69th IEEE Vehicular Technology Conference (VTC-Spring '09)*, pages 1–5, Apr. 2009.
- [102] J. T. MacDonald. A survey of spectrum occupancy in Chicago. Technical report, Illinois Institute of Technology, 2007.
- [103] S. Mangold, Z. Zhong, K. Challapali, and C.-T. Chou. Spectrum agile radio: Radio resource measurements for opportunistic spectrum usage. In *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '04)*, volume 6, pages 3467–3471, Dallas, USA, 2004.
- [104] S. Mangold, S. Shankar, and L. Berlemann. Spectrum agile radio: A society of machines with value-orientation. In *Proceedings of 11th European Wireless Conference 2005*, volume 2, pages 539–546, Nicosia, Cyprus, April 2005.
- [105] A. Martian, I. Marcu, and I. Marghescu. Spectrum occupancy in an urban environment: A cognitive radio approach. In *Proc. of the 6th Advanced International Conference on Telecommunications (AICT '10)*, pages 25–29, May 2010.
- [106] R. Matheson. Strategies for spectrum usage measurements. In *Proceedings of the IEEE International Symposium on Record Electromagnetic Compatibility, 1988*, pages 235–241, 1988.
- [107] R. McEliece and W. Stark. Channels with block interference. *IEEE Transactions on Information Theory*, 30(1):44–53, 1984.
- [108] M. McHenry, K. Steadman, A. E. Leu, and E. Melick. XG DSA radio system. In *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '08)*, pages 1–11, Oct. 2008.
- [109] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood. Chicago spectrum occupancy measurements & analysis and a long-term studies proposal. In *Proc. of the 1st International Workshop on Technology and Policy for Accessing Spectrum (TAPAS '06)*, 2006.
- [110] S. M. Mishra, D. Cabric, C. Chang, D. Willkomm, B. van Schewick, A. Wolisz, and R. W. Brodersen. A real time cognitive radio testbed for physical and link layer experiments. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 562–567, Baltimore (MD), Nov. 2005.
- [111] S. M. Mishra, A. Sahai, and R. W. Brodersen. Cooperative sensing among cognitive radios. In *Proc. of the IEEE International Conference on Communications (ICC '06)*, pages 1658–1663, June 2006.
- [112] J. Mitola III. *Cognitive Radio An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.
- [113] J. Mitola III. Cognitive radio policy languages. In *Proc. of the IEEE International Conference on Communications (ICC '09)*, pages 1–4, June 2009.
- [114] J. Mitola III. Software radios survey, critical evaluation and future directions. In *IEEE National Telesystems Conference*, pages 13/15–13/23, New York, 1992.

- [115] J. Mitola III. Cognitive radio for flexible mobile multimedia communications. In *Sixth International Workshop on Mobile Multimedia Communications (MoMuC '99)*, San Diego, CA, 1999.
- [116] J. Mitola III and G. Maguire Jr. Cognitive radio: Making software radio more personal. *IEEE Personal Communications Magazine*, 6:13–18, august 1999.
- [117] J. Mitola III and H. Man. Semantics in cognitive radio. In *Proc. of the IEEE International Conference on Semantic Computing ICSC 2009*, pages 261–266, Sept. 2009.
- [118] H. Mutlu, M. Alanyali, and D. Starobinski. Spot pricing of secondary spectrum usage in wireless cellular networks. In *Proc. of the 27th IEEE International Conference on Computer Communications (INFOCOM '08)*, pages 682–690, 2008.
- [119] C. Pantjiaros, P. Laycock, G. Gott, and S. Chan. Development of the laycock-gott occupancy model. *IEE Proceedings-Communications*, 144(1):33–39, Feb. 1997.
- [120] E. Peh and Y.-C. Liang. Optimization for cooperative sensing in cognitive radio networks. In *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pages 27–32, Mar. 2007.
- [121] E. C. Y. Peh, Y. C. Liang, Y. L. Guan, and Y. Zeng. Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view. *IEEE Transactions on Vehicular Technology*, 58(9):5294–5299, Nov. 2009.
- [122] J. M. Peha. Approaches to spectrum sharing. *IEEE Communications Magazine*, 43(2): 10–12, Feb. 2005.
- [123] D. Pemstein, K. M. Quinn, and A. D. Martin. Scythe statistical library, v. 1.0.2, 2007. URL <http://scythe.wustl.edu>.
- [124] D. Percival, M. Kraetzel, and M. Britton. A markov model for HF spectral occupancy in central australia. In *Proceedings of the 7th International Conference on HF Radio Systems and Techniques*, pages 14–18, 7–10 July 1997.
- [125] F. Perich. Policy-based network management for next generation spectrum access control. In *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pages 496–506, Apr. 2007.
- [126] F. Perich and M. McHenry. Policy-based spectrum access control for dynamic spectrum access network radios. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(1):21–27, 2009.
- [127] H. N. Pham, Y. Zhang, P. E. Engelstad, T. Skeie, and F. Eliassen. Optimal cooperative spectrum sensing in cognitive sensor networks. In *IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing*, pages 1073–1079, New York, NY, USA, 2009. ACM.
- [128] G. Pólya. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Mathematische Zeitschrift*, 8(3):171–181, 1920.
- [129] J. Proakis. *Digital Communications*. McGraw-Hill Science/Engineering/Math, 4 edition, 2000.
- [130] P. Qihang, Z. Kun, W. Jun, and L. Shaoqian. A distributed spectrum sensing scheme based on credibility and evidence theory in cognitive radio context. In *Proc. of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, pages 1–5, Sept. 2006.
- [131] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, 2001.
- [132] T. Renk, C. Kloeck, F. K. Jondral, P. Cordier, O. Holland, and F. Negredo. Spectrum

- measurements supporting reconfiguration in heterogeneous networks. In *Proc. of the 16th IST Mobile & Wireless Communications Summit*, pages 1–5, 1–5 July 2007.
- [133] D. A. Roberson, C. S. Hood, J. L. LoCicero, and J. T. MacDonald. Spectral occupancy and interference studies in support of cognitive radio technology deployment. In *Proceedings of the 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks (SDR '06)*, pages 26–35, 25–25 Sept. 2006.
- [134] K. Ruttik, K. Koufos, and R. Jantti. Detection of unknown signals in a fading environment. *IEEE Communications Letters*, 13(7):498–500, 2009.
- [135] RWTH Aachen University. Static spectrum occupancy measurement campaign. [online], 2009. URL <http://download.mobnets.rwth-aachen.de/index.php?id=102>.
- [136] A. Sahai, R. Tandra, S. M. Mishra, and N. Hoven. Fundamental design tradeoffs in cognitive radio systems. In *Proc. of the 1st International Workshop on Technology and Policy for Accessing Spectrum (TAPAS '06)*, 2006.
- [137] Shared Spectrum Company. Recent spectrum measurements, 2007. URL <http://www.sharedpectrum.com/measurements/recent.html>.
- [138] S. J. Shellhammer. Spectrum sensing in IEEE 802.22. In *In Proc. of the 1st IAPR Workshop on Cognitive Information Processing*, 2008.
- [139] C. Smith and D. Collins. *3G Wireless Networks*. McGraw-Hill, Inc., 2001.
- [140] A. Spaulding and G. Hagn. On the definition and estimation of spectrum occupancy. *IEEE Transactions on Electromagnetic Compatibility*, 19(3):269–280, 1977.
- [141] R. Steele, editor. *Mobile Radio Communications*. Pentech Press, 1992.
- [142] S. Subramani, S. Armour, D. Kaleshi, and Z. Fan. Spectrum scanning and reserve channel methods for link maintenance in cognitive radio systems. In *Proc. of the 67th IEEE Vehicular Technology Conference (VTC-Spring '08)*, pages 1944–1948, May 2008.
- [143] A. P. Subramanian, H. Gupta, S. R. Das, and M. M. Buddhikot. Fast spectrum allocation in coordinated dynamic spectrum access based cellular networks. In *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '07)*, pages 320–330, 17–20 April 2007.
- [144] C. Sun, W. Zhang, and K. B. Letaief. Cooperative spectrum sensing for cognitive radios under bandwidth constraints. In *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pages 1–5, Mar. 2007.
- [145] H. Sun, D. I. Laurenson, and C.-X. Wang. Computationally tractable model of energy detection performance over slow fading channels. *IEEE Communications Letters*, 14(10):924–926, 2010.
- [146] R. Tandra and A. Sahai. Fundamental limits on detection in low SNR under noise uncertainty. In *Proc. of the International Conference on Wireless Networks, Communications and Mobile Computing*, pages 464–469, June 2005.
- [147] R. Tandra and A. Sahai. Snr walls for signal detection. *IEEE Journal of Selected Topics in Signal Processing*, 2(1):4–17, Feb. 2008.
- [148] R. Tandra, S. M. Mishra, and A. Sahai. What is a spectrum hole and what does it take to recognize one? *Proceedings of the IEEE*, 97(5):824–848, May 2009.
- [149] J. Unnikrishnan and V. V. Veeravalli. Cooperative sensing for primary detection in cognitive radio. *IEEE Journal of Selected Topics in Signal Processing*, 2(1):18–27, Feb. 2008.
- [150] H. Urkowitz. Energy detection of unknown deterministic signals. *Proceedings of the IEEE*, 55(4):523–531, 1967.
- [151] V. Vanghi, A. Damnjanovic, and B. Vojcic. *The cdma2000 System for Mobile Communi-*

- cations: 3G Wireless Evolution*. Prentice Hall PTR, 2004.
- [152] F. E. Visser, G. J. M. Janssen, and P. Pawelczak. Multinode spectrum sensing based on energy detection for dynamic spectrum access. In *Proc. of the 67th IEEE Vehicular Technology Conference (VTC-Spring '08)*, pages 1394–1398, 2008.
 - [153] P. Wang, L. Xiao, S. Zhou, and J. Wang. Optimization of detection time for channel efficiency in cognitive radio systems. In *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pages 111–115, Mar. 11–15, 2007.
 - [154] T. Weingart, G. V. Yee, D. C. Sicker, and D. Grunwald. Implementation of a reconfiguration algorithm for cognitive radio. In *Proc. of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '07)*, pages 171–180, Aug. 2007.
 - [155] T. Weiss and F. Jondral. Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency. *IEEE Communications Magazine*, 42:S8–S14, march 2004.
 - [156] T. Weiss, J. Hillenbrand, and F. Jondral. A diversity approach for the detection of idle spectral resources in spectrum pooling systems. In *Proc. of the 48th International Science Colloquium*, Ilmenau, Germany, Sept. 2003.
 - [157] T. Weiss, J. Hillenbrand, A. Krohn, and F. Jondral. Efficient signaling of spectral resources in spectrum pooling systems. In *Proceedings of the 10th Symposium on Communications and Vehicular Technology in the Benelux*, Eindhoven, Netherlands, November 2003.
 - [158] M. Wellens and P. Mähönen. Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model. In *Proc. of the 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom '09)*, pages 1–9, Apr. 2009.
 - [159] M. Wellens and P. Mähönen. Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model. *Mobile Networks and Applications*, 15:461–474, 2010.
 - [160] M. Wellens, J. Wu, and P. Mahonen. Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio. In *Proc. of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '07)*, pages 420–427, Aug. 2007.
 - [161] M. Wellens, J. Riihijrvi, and P. Mhnen. Empirical time and frequency domain models of spectrum use. *Physical Communication; Cognitive Radio Networks: Algorithms and System Design*, 2(1-2):10–32, 2009.
 - [162] J. Wepman, R. Matheson, K. Allen, and E. Pol. Spectrum usage measurements in potential pcs frequency bands. In *Proceedings of the 1st International Conference on Universal Personal Communications, 1992 (ICUPC '92)*, pages 01.03/1–01.03/6, 1992.
 - [163] K. Wessel, M. Swigulski, A. Köpke, and D. Willkomm. MiXiM - the physical layer: An architecture overview. In *Proceeding of the 2. International Workshop on OMNeT++*, pages 1–8, Mar. 2009.
 - [164] D. Willkomm. Spectrum efficient QoS support for secondary users in cognitive radio systems. Technical Report TKN-09-008, Telecommunication Networks Group, Technische Universität Berlin, Dec. 2009.
 - [165] D. Willkomm and A. Wolisz. Efficient QoS support for secondary users in cognitive radio systems. *IEEE Wireless Communications Magazine*, 17(4):16–23, Aug. 2010.
 - [166] D. Willkomm and A. Wolisz. Is oversensitive spectrum sensing the door opener for initial cognitive radio deployments? In *Proc. of the 2nd ACM MobiHoc S3 Workshop*, Sept. 2010.

-
- [167] D. Willkomm, J. Gross, and A. Wolisz. Reliable link maintenance in cognitive radio systems. In *Proc. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pages 371–378, Baltimore (MD), Nov. 2005.
- [168] D. Willkomm, M. Bohge, D. Hollós, and J. Gross. Double hopping: A new approach for dynamic frequency hopping in cognitive radio networks. Technical Report TKN-08-001, Telecommunication Networks Group, Technische Universität Berlin, Jan. 2008.
- [169] D. Willkomm, M. Bohge, D. Hollós, J. Gross, and A. Wolisz. Double hopping: A new approach for dynamic frequency hopping in cognitive radio networks. In *Proc. of the 19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08)*, Sept. 2008.
- [170] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary users in cellular networks: A large-scale measurement study. In *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '08)*, Chicago, IL, Oct. 2008. Winner of best paper award.
- [171] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary user behavior in cellular networks and implications for dynamic spectrum access. *IEEE Communications Magazine*, 47(3):88–95, Mar. 2009.
- [172] D. Willkomm, W. Hu, D. Hollós, J. Gross, and A. Wolisz. On centralized and distributed frequency assignment in cognitive radio based frequency hopping cellular networks. In *In Proc. of the 3rd International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART '10)*, Nov. 2010. Invited paper.
- [173] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. The problem of sensing unused cellular spectrum. In *Proc. of the 10th International IFIP TC 6 Networking Conference*, Lecture Notes in Computer Science, May 2011.
- [174] D. Xu and X. Liu. Opportunistic spectrum access in cognitive radio networks: when to turn off the spectrum sensors. In *Proceedings of the 4th Annual International Conference on Wireless Internet (WICON 2008)*, pages 1–6, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [175] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht. DRiVE-ing to the internet: Dynamic radio for IP services in vehicular environments. In *Proceedings of the 25th Annual IEEE Conference on Local Computer Networks LCN 2000*, pages 281 – 289, nov 2000.
- [176] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Communications Surveys Tutorials*, 11(1):116–130, 2009.