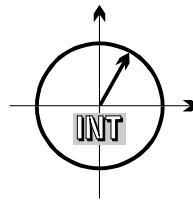


■ *Forschungsberichte aus dem
Institut für Nachrichtentechnik
der Universität Karlsruhe (TH)*



Ulrich Berthold

■ **Dynamic Spectrum Access Using OFDM-based Overlay Systems**

■ Band 21

Copyright: Institut für Nachrichtentechnik
Universität Karlsruhe (TH), 2009

Druck: E&B printware, Käpplestr. 10
76131 Karlsruhe, Tel. 0721/962 26

ISSN: 1433-3821

**Forschungsberichte aus dem Institut für Nachrichtentechnik der
Universität Karlsruhe (TH)**

Herausgeber: Prof. Dr. rer. nat. Friedrich Jondral

- Band 1 Marcel Kohl
Simulationsmodelle für die Bewertung von Satellitenübertragungsstrecken im 20/30 GHz Bereich
- Band 2 Christoph Delfs
Zeit-Frequenz-Signalanalyse: Lineare und quadratische Verfahren sowie vergleichende Untersuchungen zur Klassifikation von Klaviertönen
- Band 3 Gunnar Wetzker
Maximum-Likelihood Akquisition von Direct Sequence Spread-Spectrum Signalen
- Band 4 Anne Wiesler
Parametergesteuertes Software Radio für Mobilfunksysteme
- Band 5 Karl Lütjen
Systeme und Verfahren für strukturelle Musteranalysen mit Produktionsnetzen
- Band 6 Ralf Machauer
Multicode-Detektion im UMTS
- Band 7 Gunther M. A. Sessler
Schnell konvergierender Polynomial Expansion Multiuser Detektor mit niedriger Komplexität
- Band 8 Henrik Schober
Breitbandige OFDM Funkübertragung bei hohen Teilnehmergeschwindigkeiten

**Forschungsberichte aus dem Institut für Nachrichtentechnik der
Universität Karlsruhe (TH)**

Herausgeber: Prof. Dr. rer. nat. Friedrich Jondral

- Band 9 Arnd-Ragnar Rhiemeier
Modulares Software Defined Radio
- Band 10 Mustafa Mengüç Öner
Air Interface Identification for Software Radio Systems
- Band 11 Fatih Çapar
**Dynamische Spektrumverwaltung und elektronische Echtzeitver-
marktung von Funkspektren in Hotspotnetzen**
- Band 12 Ihan Martoyo
Frequency Domain Equalization in CDMA Detection
- Band 13 Timo Weiß
OFDM-basiertes Spectrum Pooling
- Band 14 Wojciech Kuropatwiński-Kaiser
**Messung von MIMO-Kapazitäten mit einem auf GSM-
Komponenten basierenden Demonstrator**
- Band 15 Piotr Rykaczewski
**Kompensierung hochfrequenzbedingter I/Q-Fehler in Software
Defined Radio**
- Band 16 Michael Eisenacher
**Kompensierung der stochastischen Eigenschaften von Ultra-
Wideband-Signalen**
- Band 17 Clemens Klöck
Auction-based Medium Access Control

**Forschungsberichte aus dem Institut für Nachrichtentechnik der
Universität Karlsruhe (TH)**

Herausgeber: Prof. Dr. rer. nat. Friedrich Jondral

- Band 18 Martin Henkel
**Architektur eines DRM-Empfängers und Basisbandalgorithmen
zur Frequenzakquisition und Kanalschätzung**
- Band 19 Stefan Edinger
**Mehrträgerverfahren mit dynamisch-adaptiver Modulation zur
unterbrechungsfreien Datenübertragung in Störfällen**
- Band 20 Volker Blaschke
Multiband Cognitive Radio-Systeme
- Band 21 Ulrich Berthold
Dynamic Spectrum Access Using OFDM-based Overlay Systems

Vorwort des Herausgebers

Die *Effizienz der Spektrumsnutzung*¹ für ein innerhalb einer festen Bandbreite im Einsatz befindliches System wird gemessen als die mittlere Anzahl von Bits, die pro Sekunde und pro Hertz in einer bestimmten Fläche übertragen wird; ihre Einheit ist damit bit/s/Hz/m^2 . In den vergangenen Jahren durchgeführte Messungen, die eine Spektrumseffizienz von maximal fünfzehn bis zwanzig Prozent ergeben haben, führen zu der berechtigten Annahme, dass dynamische Zugriffsverfahren auf die Übertragungsressource zumindest dabei helfen können, deren Knappheit zu reduzieren. Die Anwendung solcher Verfahren wird im Wesentlichen durch den Fortschritt von Mikroelektronik und Nachrichtentechnik möglich, sie bedarf jedoch auch der Berücksichtigung wirtschaftlicher, regulatorischer und damit nicht zuletzt politischer Randbedingungen. Letztlich wird diese Entwicklung dazu führen, dass in größeren Bereichen des elektromagnetischen Spektrums von der heute noch oft praktizierten starren Zuteilung von Übertragungsressourcen zu einem dynamischen Spektrumszugriff übergegangen wird.

Die Dissertation von Ulrich Berthold behandelt in diesem Zusammenhang auftretende technische Aufgabenstellungen. Zunächst einmal geht es darum, eine Übersicht über die verschiedenen Aspekte und Ansätze des dynamischen Spektrumszugriffs zu geben. Da der Übergang zu einem dynamischen Spektrumszugriff langsam erfolgen wird, werden in der einschlägigen Literatur seit einiger Zeit Möglichkeiten diskutiert, die innerhalb der Spektren heute existierender Systeme brach liegenden Ressourcen für Overlay-Systeme zu verwenden. Dabei ist darauf zu achten, dass erstens das primäre System von dem Overlay-System nicht gestört wird und zweitens keine Veränderungen am Primärsystem notwendig werden. Sämtliche Algorithmen, die die Koexistenz der Systeme im selben Frequenzbereich gestatten, sind also im Overlay-System zu realisieren.

Orthogonal Frequency Division Multiplexing (OFDM) bietet sich aufgrund seiner flexiblen Mehrträgerstruktur als Grundlage der Übertragung in Overlay-Systemen an. OFDM-basierte Overlay-Systeme mit Sterntopologie werden am Institut für Nachrichtentechnik der Universität Karlsruhe (TH) seit mehreren Jahren untersucht². Die Aufgabe der Arbeit von Ulrich Berthold besteht nun in der Untersuchung von Overlay-Systemen, die im Ad-Hoc-Mode betrieben werden, bei denen es also keine Sonderstellung einer bestimmten Station und damit keine zentrale

¹Die *Effizienz der Spektrumsnutzung* wird vereinfachend als *Spektrumseffizienz* bezeichnet.

²Siehe z.B.: Timo Weiß: *OFDM-basiertes Spectrum Pooling*. Dissertation, Forschungsberichte aus dem Institut für Nachrichtentechnik der Universität Karlsruhe (TH), Band 13, Karlsruhe 2004.

Steuerung gibt. Dabei ist die vorliegende Dissertation im Wesentlichen der Untersuchung von Medium Access Control (MAC) Layer Aspekten gewidmet.

Die wesentlichen Beiträge von Herrn Berthold zum Fortschritt von Wissenschaft und Technik sind:

- Die Ableitung und die Bewertung eines für in Ad-Hoc-Overlay-Systemen anwendbaren Detektionsalgorithmus für Primärnutzersignale
- Die Bestimmung theoretischer Grenzen für die minimale Anzahl von Overlay-Stationen, die zur sicheren Entdeckung eines Primärnutzersignals notwendig sind
- Die Ableitung von Verteilungen für Stationen im Ad-Hoc-System, die Verbindungen von jedem zu jedem anderen Teilnehmer erfordern, aus der Spieltheorie
- Die Definition eines Zugriffsprotokolls für im Ad-Hoc-Mode arbeitende Overlay-Systeme

Es bleibt zu wünschen, dass die vorliegende Arbeit die Diskussionen, die zu einer effizienteren Nutzung der elektromagnetischen Übertragungsressource führen, nachhaltig beleben wird.

Karlsruhe, im Februar 2009
Friedrich Jondral

Dynamic Spectrum Access Using OFDM-based Overlay Systems

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

der Fakultät für
Elektrotechnik und Informationstechnik
der Universität Fridericiana Karlsruhe

genehmigte

DISSERTATION

von

Dipl.-Ing. Ulrich Berthold

aus

Mainz

Tag der mündlichen Prüfung:

05.02.2009

Hauptreferent:

Prof. Dr. rer. nat. Friedrich Jondral

Korreferent:

Prof. Dr. Mihaela van der Schaar

Korreferent:

Dr.-Ing. Michael Schnell

Acknowledgement

This dissertation results from my work at the Institut für Nachrichtentechnik (INT) at the Universität Karlsruhe (TH) and could not have been finished without the support of several people. First of all, I would like to sincerely thank Prof. Dr. rer.nat. Friedrich K. Jondral for the supervision of my work and the opportunity to accomplish my dissertation. During the time at the INT I could learn a lot from his experience. I also thank him for undertaking the task of the first reviewer.

Furthermore, I thank Prof. Dr. Mihaela van der Schaar for inviting me to join her research group at the Multimedia Communications and Systems Laboratory at the University of California, Los Angeles (UCLA) for three months. It was a very good and productive time with many helpful discussions. But especially I want to thank her for spending her time and effort for the task of the second reviewer. Many thanks also to Dr.-Ing. Michael Schnell who also provided a review of my dissertation.

I was happy to work with Dipl.-Ing. Sinja Brandes within the projects TAKOKO and OOS and would like to thank her for the cooperation and valuable discussions. Many thanks also to my colleagues at the INT for the open-minded and discussion-friendly atmosphere. Especially, I would like to express my gratitude to Dipl.-Ing. Dennis Burgkhardt and Dr.-Ing. Holger Jäkel for proofreading my dissertation. Furthermore, I thank our secretaries Gabriele Kuntermann and Beate Mast, our designer Angelika Olbrich and system administrator Dipl.-Ing. (FH) Reiner Linnenkohl who made daily routine a lot easier.

I also want to thank Friedemann Kalmbach for being a great mentor during the past years and for always helping me to see the big picture. Many thanks goes to my parents and brothers who were always there and supported me when I needed them. My utmost gratitude belongs to my Lord Jesus Christ who gives meaning and hope to life.

Zusammenfassung

Mit einem wachsenden Markt für Mobilkommunikation nimmt der Bedarf an dafür geeignetem Spektrum zu, das von technischer Seite gesehen die Hauptressource für drahtlose Datenübertragung darstellt. Bedingt durch die physikalischen Ausbreitungseigenschaften von elektromagnetischen Wellen ist diese Ressource nur begrenzt verfügbar, womit eine effiziente Nutzung immer wichtiger wird. Die bisher übliche Vorgehensweise der Regulierungsbehörden bei der Frequenzvergabe besteht darin, den einzelnen Diensten und Diensteanbietern die nötigen Frequenzbereiche für eine exklusive Nutzung zuzuteilen. Dies war ein sinnvoller Ansatz für die letzten Jahrzehnte. Aufgrund aktueller technologischer Entwicklungen im Bereich der Kommunikationshardware, die den Entwurf von kleinen und gleichzeitig kostengünstigen Mobilgeräten mit hoher Leistungsfähigkeit und Datenrate ermöglichen, nimmt die Anzahl der mobilen Dienste jedoch nachhaltig zu. Dies führt dazu, dass es für Diensteanbieter sehr schwierig ist, neue Dienste in Betrieb zu nehmen oder bereits vorhandene zu erweitern. Messungen zeigen jedoch, dass nur ein kleiner Teil der zugewiesenen Frequenzbereiche über die Zeit gesehen auch tatsächlich genutzt wird. Diese Diskrepanz zwischen zugeteilten und tatsächlich genutzten Frequenzbereichen hat in den letzten Jahren zu einem weltweiten Forschungsschwerpunkt im Bereich des dynamischen Spektrumzugriffs geführt, mit dem Ziel, der scheinbaren Knappheit der Ressourcen mit ihrer effizienten Nutzung zu begegnen. In diesem Zusammenhang müssen regulatorische, wirtschaftliche und technische Aspekte berücksichtigt werden. Auf der einen Seite muss die Zuteilung der Frequenzen flexibler gestaltet werden, um sie dem sich ändernden Bedarf an Funkressourcen der Dienste zeitnah anpassen zu können. Zum anderen muss aber auch die verwendete Technologie flexibler werden, d. h. die Kommunikationssysteme müssen sich schnell an die vorhandenen Bedingungen anpassen können, um die dynamisch entstehenden Frequenzlücken auch nutzen zu können. Hierbei spielt der Bereich "Cognitive Radio" eine wesentliche Rolle, da Cognitive Radios die notwendigen technischen Voraussetzungen für einen dynamischen Spektrumzugriff zur Verfügung stellen.

Die Untergruppe der Overlay-Systeme stellt dabei einen vielversprechenden Ansatz zur Überwindung der scheinbaren Ressourcenknappheit dar, insbesondere in der Übergangsphase zu einem komplett dynamischen Spektrumzugriff. Die Idee besteht darin, die ungenutzten Bereiche des Spektrums für den Betrieb eines zusätzlichen, unabhängigen Systems im selben Frequenzbereich zu nutzen. Overlay-Systeme haben dabei den Vorteil, dass sie relativ einfach in bereits lizenzierten

Frequenzbändern in Betrieb genommen werden können, ohne dass dabei die vorhandene Infrastruktur der lizenzierten Systeme modifiziert werden muss. Dies bedeutet im Gegenzug, dass die gesamte zusätzliche Signalverarbeitung, die durch die Koexistenz notwendig wird, im Overlay-System implementiert werden muss. Es dürfen nur die freien Ressourcen, die momentan nicht durch das lizenzierte System belegt sind, vom Overlay-System benutzt werden. Dies führt zu einem signifikanten Einfluss auf das Systemdesign, sowohl in der Medienzugriffsschicht als auch in der physikalischen Schicht.

In dieser Arbeit wird eine Übersicht der verschiedenen Aspekte und Ansätze des dynamischen Spektrumzugriffs gegeben und dieser durch die jeweilige Einordnung der Ansätze in verschiedene Kategorien strukturiert. Der Schwerpunkt liegt dabei auf OFDM-basierten Overlay-Systemen. In diesem Zusammenhang werden charakteristische Merkmale von lizenzierten Systemen diskutiert sowie Richtlinien zum Entwurf und zur Parametrisierung entsprechender Overlay-Systeme gegeben. Ein wesentlicher Beitrag dieser Arbeit besteht in einem Ansatz zur verteilten Detektion von lizenzierten Systemen für Overlay-Systeme, die im Ad-Hoc Mode betrieben werden. Anhand eines Modells aus der Graphentheorie werden theoretische Schranken für die minimal notwendige Knotendichte hergeleitet, die eine geforderte mittlere Netzwerk-Entdeckungswahrscheinlichkeit sicherstellen. Weiterhin müssen die Detektionsphasen aller beteiligten Overlay-Stationen zeitlich synchronisiert werden, damit Sendesignale von Overlay-Stationen die Detektion des lizenzierten Systems nicht beeinträchtigen. Dies kann durch das vorgeschlagene AHOMAC-(Medienzugriff für Overlay-Systeme im Ad-Hoc Mode)-Protokoll erreicht werden. Einen weiteren wichtigen Aspekt dieser Arbeit stellt die Detektion von spektralen Ressourcen in einer Multiband-Umgebung dar. Hierbei wird der insgesamt für einen Overlay-Betrieb zur Verfügung stehende Frequenzbereich in mehrere Teilbänder unterteilt. Aufgrund seiner beschränkten Bandbreite kann das Overlay-System jedoch nur in jeweils einem der Teilbänder betrieben werden. Die Teilbänder weisen dabei eine unterschiedliche mittlere Belegung durch das lizenzierte System auf, sodass das Overlay-System jeweils das aktuell beste Teilband hinsichtlich der freien Ressourcen finden muss. Diese Problemstellung kann als Lernaufgabe betrachtet und als Markov'scher Entscheidungsprozess modelliert werden. Mit Hilfe von Methoden des bestärkenden Lernens wird ein Lernalgorithmus entwickelt, der ein ausgeglichenes Verhältnis zwischen dem Erforschen der Belegungssituation in den einzelnen Teilbändern und dem Nutzen der bereits vorhandenen Information zur Übertragung von Daten erreicht.

Abstract

With a steadily growing market for mobile communications the demand for its main technological resource, namely suitable spectrum, is increasing. Since this resource is naturally limited by its physical properties, an efficient spectrum use gains importance. The current policy of the regulatory body is to assign defined parts of the spectrum exclusively to distinguished services and service providers. This has been a practical approach in the last century. However, due to the recent technological developments in communication hardware allowing the design of small mobile devices capable of processing and transmitting high data rates at affordable costs, many new services are emerging. In consequence, it is very difficult for service providers to obtain additional frequency bands and thus to establish new services or extend existing ones. Nevertheless, measurements show that only a small percentage of the assigned spectrum is actually used. This discrepancy has recently put the focus of research on increasing the efficiency in spectrum use, where dynamic spectrum access is an important step towards this goal. This involves regulatory and economical as well as technological aspects. On the one hand, regulations must become more flexible to meet the changing demand for resources of the various services. On the other hand, technology must become more flexible, that is, reconfigurable and spectrum agile. Here, the area of cognitive radio plays an important role as an enabling technology for dynamic spectrum access.

Within this area the concept of overlay systems represents one approach to overcome the seemingly spectrum scarcity, suitable especially in the transition phase to fully applied dynamic spectrum access. The idea is to exploit unused spectrum parts by operating an additional system in the same frequency band. Overlay systems have the advantage that they can easily be deployed in already licensed frequency bands without the need for changing existing infrastructure. Nevertheless, this means that all additional signal processing required for realizing the coexistence must be implemented in the overlay system. Only the free spectral resources not needed by the primary system may be used by the overlay system, leading to a significant impact on the design of the medium access control layer as well as the physical layer.

In this work, various aspects and approaches of dynamic spectrum access are reviewed and structured in different categories. The focus is on overlay systems based on orthogonal frequency division multiplex (OFDM) including a discussion of characteristic parameters of primary systems and guidelines for designing OFDM-based overlay systems. One major contribution of this work is the distributed detection of primary users for overlay systems operating in ad hoc mode.

Using a model from graph theory, theoretical bounds for the minimum necessary node density for still achieving a required network detection probability are derived. Furthermore, the detection periods of all contributing secondary users have to be synchronized to ensure a reliable detection of the primary user signal. This can be achieved by the proposed ad hoc overlay medium access control (AHOMAC) protocol. The second main focus of this work is the detection of spectral resources in a multiband environment. A multiband scenario assumes several frequency bands, each with different occupation characteristics, and an overlay system that can only operate in one frequency band simultaneously. Therefore, the overlay system has to find the frequency band providing the best opportunity for transmitting data. This problem can be interpreted as a learning problem and is modeled as Markov decision process. With the help of reinforcement learning a corresponding learning algorithm is developed, giving a balance between exploring the status regarding available spectrum of all frequency bands and exploiting the already detected resources.

Content

1	Introduction	1
1.1	Motivation	1
1.2	Outline of the Work	3
2	Dynamic Spectrum Access	5
2.1	Increasing the Efficiency of Spectrum Access	5
2.2	Categorizing Spectrum Access	7
2.3	Advances in Dynamic Spectrum Access	9
2.3.1	Regulatory Aspects	9
2.3.2	Technological Aspects	9
2.3.3	Standardization	11
3	OFDM based Overlay Systems	12
3.1	System Model	12
3.1.1	Overlay Scenarios	14
3.1.2	Characteristic Parameters of Primary User Systems	19
3.1.3	Design Parameters of OFDM Secondary User Systems	20
3.2	Detection Subsystem	22
3.2.1	The Detector Model	24
3.3	Traffic Model for Primary User Systems	27
3.3.1	Simple Model	27
3.3.2	Model for a TDMA Based Primary User System	28
3.4	Optimizing the Secondary User System's Parameter Configuration	30
3.4.1	Duration of the Detection Phase	31
3.4.2	Update Interval	34
3.5	Multiple Spectrum Occupation Zones	36
3.5.1	Maximum Allowable Transmission Range of Secondary Users	37

3.5.2	Secondary User Systems Operating in Multiple Spectrum Occupation Zones	38
4	Overlay Systems in Ad Hoc Mode	41
4.1	Ad Hoc Networks	41
4.2	Model Based on Graph Theory	43
4.2.1	Number of Neighbors	44
4.3	Distributed Detection in Ad Hoc Networks	47
4.3.1	Distributed Detection	48
4.3.2	Theoretical Bounds	52
4.3.3	Efficient Signaling of Detection Results	59
4.4	Medium Access Control for Overlay Systems in Ad Hoc Mode . .	73
4.4.1	Coordination of Spectrum Occupancy Measurements . . .	74
5	Detection of Spectral Resources in a Multiband Environment	80
5.1	Reinforcement Learning	82
5.1.1	Formal Model of the Learning Problem	84
5.1.2	Dynamic Programming	88
5.1.3	Monte Carlo Methods	92
5.1.4	Temporal-Difference Learning	94
5.2	Problem Formulation as Markov Decision Process	98
5.2.1	Reference Model	99
5.2.2	Basic Model	101
5.2.3	Learning Strategy for Basic Model	106
5.2.4	Extended Model with Out-of-Band Detection	110
5.2.5	Modified Learning Strategy for the Extended Model . . .	111
6	Conclusion	116
	Acronyms	119
	Bibliography	122

Theses	128
Index	129
Förderung	131
Curriculum Vitae	133

1 Introduction

In the last few decades society began an impressive transformation process towards an information society. Similar to the industrialization in the late 19th century that ushered in a completely new era, the current transformation process also impacts all areas of society in a dramatical way by increasing the speed, amount, and availability of information [70]. With the growing possibilities offered by the telecommunication industry also the communication profile of the user in general changes, expecting instant access to information and communication services anywhere at any time. Recent developments in communication hardware allow the design of small mobile devices which are capable of processing and transmitting high bit rate data at affordable costs, resulting in a growing mass market for mobile multimedia communications, including for example digital television, music and video download or mobile access to navigation services. Meanwhile, mobile communication services are taken for granted by the user, leading to a growing demand for further services with even higher data rates. This in turn further accelerates the speed of technological developments, resulting in the need for highly efficient communication technologies. Especially in the wireless and mobile domain this raises new challenges that have to be handled to be able meet the growing number of users as well as the growing “hunger” for a high data rate of each individual user.

1.1 Motivation

The main technological resource for providing mobile multimedia services is the available spectrum suitable for wireless communications which represents a pillar for mobile communication. This resource is limited due to the physical properties of electro-magnetical wave propagation. When looking at the frequency assignments managed by the regulatory body, for example the Federal Communications Commission (FCC) in the United States, and prices recently paid in frequency band auctions, it is very difficult and expensive for service providers to obtain new frequency bands for deploying new wireless communication systems or extending existing ones. Nevertheless, measurements show [55, 56] that although nearly all frequency bands are assigned, most of them are not used in a very efficient way, resulting in a very fragmented spectrum occupation in the time-frequency plane as depicted in Figure 1.1. Unused parts of the time-frequency plane are also

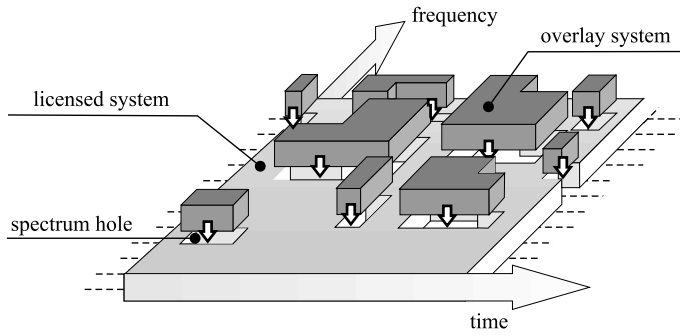


Figure 1.1 Spectrum access of the PU system in the time/frequency plane and the SU system filling the spectrum holes.

called spectrum holes. To increase the efficiency in spectrum use, there are several approaches. In cooperative systems the unused spectrum parts are used by the same provider, but for a different service, as for example in general packet radio service (GPRS), which uses available slots for the transmission of data if a Global System for Mobile communication (GSM) system is not working to full capacity. Another concept are underlay systems which comprise for example ultra-wideband (UWB) systems [62], used for short-range high-bandwidth communications. These systems coexist in the same frequency bands assigned to licensed systems and therefore accept interference in a certain extent. Nevertheless, the coexistence is enabled by spreading the signal over a very large bandwidth (up to several GHz) and using a power level below the noise level.

A different approach is dynamic spectrum access, in which available spectrum is automatically and dynamically assigned to different services while trying to minimize the interference between different systems. One approach within dynamic spectrum access (DSA) is the concept of overlay systems. A given frequency band is licensed to a primary user (PU) system and the occurring spectrum holes are exploited by an overlay system also called secondary user (SU) system, as illustrated in Figure 1.1. Both systems coexist in the same frequency band. The PU system must not be influenced in its performance, leading to several challenges in the design of the SU system and especially its detection subsystem, which is the focus of this work.

1.2 Outline of the Work

The research community in the area of dynamic spectrum access has been steadily growing in the last years, resulting in a variety of different approaches and concepts for increasing the efficiency in spectrum use. Not only with the International Symposium on Dynamic Spectrum Access Networks (DySPAN) a new series of major conferences of the Institute of Electrical and Electronics Engineers (IEEE) Communications Society (ComSoc) dedicated to dynamic spectrum access was established in 2005, but nearly all other major conferences meanwhile started their own tracks in this area. Due to the manifold issues and approaches in the area of dynamic spectrum access, in Chapter 2 an overview is given, presenting the main threads of current research. For this, a structure enabling a coarse categorization of the presented approaches is proposed followed by a brief discussion of current efforts in standardization.

The focus of this work is on overlay systems based on orthogonal frequency division multiplex (OFDM), and therefore in Chapter 3 the concept of OFDM-based overlay systems is discussed. After a short introduction to the basics of OFDM the system model for the coexistence of two independent systems operating in the same frequency band is presented, including three overlay scenarios serving as examples for possible applications. The illuminated examples appear to have different assumptions and characteristics, which leads to the discussion of general characteristic parameters of PU systems relevant to an overlay context and in a corresponding way to general guidelines for designing OFDM-based overlay systems. Furthermore, the applied detector model is introduced, based on which the configuration of the SU system parameters is optimized. When considering multiple occupation zones, that is, multiple neighboring areas with a different spectrum occupation, an SU system with stations in both areas may only use subchannels that are not occupied in both zones. The available capacity in this scenario is discussed for two occupation zones based on a simple traffic model for the spectrum access of the PU system.

Chapter 4 is devoted to overlay systems operating in ad hoc mode. After a short introduction to ad hoc networks the basics of graph theory providing a formal model for ad hoc networks are summarized. In combination with the detector model introduced in Chapter 3 the network detection probability is defined and theoretical bounds for the minimum number of nodes necessary for achieving a given network detection probability are derived. Thereby, a scenario with an infinite system area and constant node density as well as a scenario with a finite system area and therefore with effects at the fringe, are investigated. In the next step, a novel mechanism

for signaling the detection results within an ad hoc network is developed. The synchronization of the detection phases of all involved SUs is vital for SU systems, since otherwise SU transmissions would influence the detection of PU system signals. Therefore, an approach for the coordination of detection phases in ad hoc overlay systems is proposed, which is part of the medium access layer.

In Chapter 5 a multiband scenario is considered, where the bandwidth of the PU system is larger than the bandwidth of the SU system. Thus, there are several frequency bands available, but the SU system can only operate in one frequency band simultaneously. When resources run low in the current frequency band, there is a chance that in another frequency band the conditions are better for the SU system and that switching frequency bands is a reasonable option to enable a continued transmission of data. However, since the SU system can only be tuned into one frequency band at a time, it has to find a good trade-off between exploitation and exploration. That is, between playing safe and staying in the current frequency band where it knows how much data it can send, and switching to other frequency bands with the chance of finding one with more resources, but also with the risk of switching to a band with less resources. This problem can be interpreted as a learning problem. Therefore, a model is developed, representing the detection spectral resources as a Markov decision process (MDP) which can be solved by methods from reinforcement learning. Finally, this work is concluded in Chapter 6.

2 Dynamic Spectrum Access

Motivated by a seeming scarcity of available spectrum suitable for wireless communications, research in the area of efficient spectrum access has grown significantly in recent years. Measurements show that only a small percentage of the licensed spectrum actually is used, leaving enough potential for increasing the efficiency in spectrum use. There are different levels in each of which the efficiency can be increased and also the different approaches of dynamic spectrum access can be structured by using different categories. Furthermore, for successfully moving towards a general dynamic spectrum access, not only technological but also regulatory and economical aspects have to be taken into account.

2.1 Increasing the Efficiency of Spectrum Access

There exist mainly three different levels coming into consideration for increasing the efficiency in spectrum use, as depicted in Figure 2.1. These levels can be distinguished by asking either *how* to access the spectrum or *when and where* to access it regarding the frequency range. Both aspects complement each other since they are related to different layers and thus can be combined. The question how to access the spectrum is an issue of a point-to-point transmission and is therefore handled at the physical (PHY) layer. In contrast, the medium access control (MAC) layer handles the question when and where to access on a network level, that is, it coordinates the spectrum access of various users within a specific network. However, these levels for increasing the efficiency in spectrum access exist in every conventional communication system. The novel idea of dynamic spectrum access is the third level, which controls the spectrum access on an inter-network level. Generally speaking, it ensures an efficient coexistence of several wireless systems in a given frequency range, which can include very large parts of the spectrum. The dynamic spectrum access component is an independent entity but is tightly connected to the first two levels, since it has significant influence on the MAC and PHY layer.

Efficiency of a Point-to-Point Transmission

Increasing the efficiency of spectrum access in a point-to-point transmission includes all aspects on how to access the medium, under the prerequisite that the “when and where” is given. Accordingly, these optimizations are mainly per-

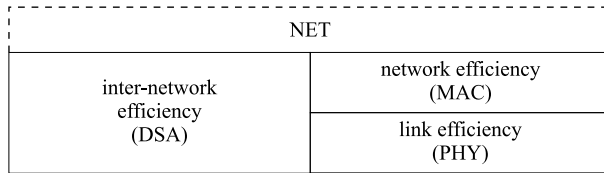


Figure 2.1 Different types of efficiency that contribute to an overall increased efficiency in spectrum use.

formed in the physical layer. There exists a variety of different concepts in this area, contributing to a higher spectral efficiency, including OFDM [58], multiple input multiple output (MIMO) [77], or adaptive modulation [45]. All these concepts follow the objective of transmitting as much information as possible per time and frequency. One example is the transition from analog to digital television (TV) broadcast: Instead of transmitting a single TV program, it is now possible to transmit two or more program channels within the same bandwidth. The efficiency in the physical layer is also referred to as spectral efficiency.

Efficiency of a Communication Network

On a node level efficiency can be increased by optimizing the spectrum access of the participating nodes in terms of an overall network throughput. This takes place in the MAC layer and includes the avoidance of collisions and minimization of signaling and control information. A simple example illustrating the potential of an optimization on the network layer is ALOHA. Pure ALOHA only achieves a throughput of 18 percent. The introduction of slots representing a different access rule double the throughput, and therefore the efficiency of spectrum access to 36 percent [7].

Efficiency of Multiple Coexisting Communication Networks

Even if the medium access of several nodes within a network is optimized, from a global perspective this only results in a high efficiency of spectrum access when the network is operating under full load. This is not the case in most of the systems, resulting in “blocked spectrum,” that is, unused, but reserved parts of the spectrum. Hence, these parts are blocked for other potential users.

Therefore, from a more global point of view, another possibility for increasing the overall efficiency of spectrum access is to optimize its coordination on an inter-network level. This is equivalent to avoiding blocked spectrum. The demand for

spectrum is time variant and therefore dynamic. Blocked spectrum originates from a spectrum allocation strategy which also may be dynamic, but does not perfectly match the demand. In order to avoid blocked spectrum, the allocation must be as dynamic as the demand. Note that all following considerations are only feasible as long as there is overall enough (blocked or unblocked) spectrum available, that is, meeting the demand for spectrum is only an issue of the allocation strategy and not of its general availability. Optimizing the strategies for spectrum access and constantly adapting it to the current demand leads to the concept of dynamic spectrum access. This actually means a different view on regulation—licenses must become much more dynamic and have a finer granularity regarding bandwidth, time, and location. But not only the changed view on regulation is necessary. This required flexibility can only be achieved by tightly coupling policy, technological, and economical aspects.

2.2 Categorizing Spectrum Access

Using a very simplified view, the electromagnetic spectrum is accessed by a variety of different transmitters and receivers designed for miscellaneous purposes, for example communication systems or sensor networks. In order to enable a successful coexistence, the spectrum access has to be coordinated, that is, each transmitter needs to know exactly when and in which frequency range it is allowed to transmit. In the same manner, the receiver or receivers paired to that transmitter have to be informed at which frequency and with which bandwidth and modulation mode they may expect the respective transmission. In general, several receivers and transmitters are related to each other, forming a network that is supporting a certain standard. Due to physical limitations and hardware constraints, nodes within a specific network should be operating in a similar frequency range.

Based on these basic observations, spectrum access can be structured in a matrix as shown in Figure 2.2. There are two main levels for the coordination of spectrum access. On the network level, only a coarse coordination of spectrum regions is performed, for example by regulatory bodies assigning parts of the spectrum to services and operators. Within this given frequency block, spectrum access regarding the involved nodes is coordinated by the applied standard. In the process of determining frequency and time of a node's transmission, regulation is responsible for the coarse coordination (network level) and the standard then takes care of the detailed access coordination between participating nodes (node level). On the other hand, independent of the level, spectrum access can be divided into centralized and

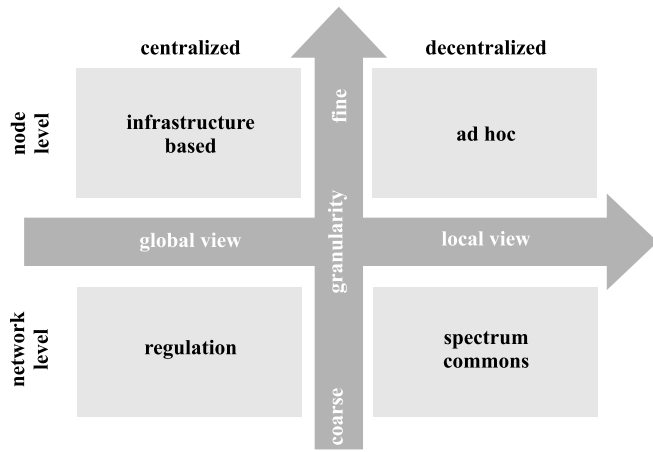


Figure 2.2 Different approaches and structure for categorizing spectrum access.

decentralized approaches, using either a global or local view, respectively. Regarding the network level, *regulation* embodies the centralized approach. A central instance, here the regulatory body, is in total control of all spectrum assignments. In case of a decentralized approach, we have the opposite situation: Several independent standards coexist in the same frequency region without any prioritization, as for example in the industrial, scientific and medical (ISM) bands. This is referred to as *spectrum commons* [87]. Note, that it is only important whether there is a central coordination instance or not. In [87] the term “dynamic exclusive use” is coined to establish a third category, grouping concepts that propose local dynamic spectrum assignment, for example [85], and spectrum property rights [48]. Nevertheless, these concepts have in common with regulation the time and location dependency, and also especially the central coordination instance. In fact, regulation varies frequency assignments in a much larger time scale and for larger geographical regions, but nevertheless it is not static. Therefore, no distinction is made in the proposed categorization. On the node level, *infrastructure based networks* represent the centralized approach, since they rely on central base stations acting as masters in their dedicated cells and frequency bands. Accordingly, *ad hoc networks* stand for the distributed and decentralized approach for the coordination of spectrum access. Note, that there also are hybrid spectrum access strategies on both, the network and node level: For example, looking at the ISM bands, a dedicated frequency band is defined by regulation (centralized component), which is

then used as spectrum commons (decentralized component). On node level, a base station can be connected with nodes in the cell, which are also connected to nodes outside the cell, using ad hoc mode.

2.3 Advances in Dynamic Spectrum Access

Recent research in the area of dynamic spectrum access spans technological, regulatory and economical aspects, since a transition to a flexible regulation goes hand in hand with technological advances. The focus of this work is on the technology side, but due to the impact on regulatory issues here the the regulatory as well as technological aspects are briefly discussed and an overview on current standardization efforts is given.

2.3.1 Regulatory Aspects

To achieve a greater flexibility in regulation allowing a faster allocation of licenses with a finer granularity, the use of technical systems enabling an automated licensing of frequencies is unavoidable. Such a system only needs some basic directives from the regulation authority and then determines the best configuration of licenses according to the currently incoming requests of different systems. For an automated exchange and distribution of policies they must be described in a formal way. Therefore, policy description languages have been proposed and discussed recently, for example in [12, 39, 54]. With the help of policy description languages it now also possible to implement time-limited spectrum leasing [33, 34] or to apply concepts, where a license owner can temporarily lease not needed licenses to other users resulting in a secondary spectrum market [64].

2.3.2 Technological Aspects

The contribution of regulation to an increased efficiency in spectrum use only has an effect if the applied technology is able to implement the required policies. In other words, regulation can be only as dynamic as as the underlying technology is dynamic. Basically, from the technology side there are two main areas that are important for a dynamic spectrum access. In general, the available resources have to be detected first, before they can then be allocated to different services. This applies especially to overlay systems, but also in opportunistic spectrum access a system first needs to find a suitable space where it then can start its operation. On

the other hand, when several equally prioritized systems or users with a different but time variant demand share the same resources, the available spectrum has to be allocated in a fair manner.

Spectrum Sensing

Since detection is a vital component for cognitive radios and dynamic spectrum access, there has recently been done a lot of research in this area. A variety of different aspects have been studied which can be distinguished in several ways. Depending on the available information about the systems that are to be detected, either a simple energy detection can be applied. Alternatively or additionally known features can be extracted by the more complex feature detection, for example by analyzing cyclostationary properties of the received signals [60, 61]. Both approaches can be further enhanced when several stations collaborate and apply cooperative sensing approaches, as for example discussed in [21, 57, 59, 63, 69]. Another challenge for spectrum sensing is the large bandwidth that has to be scanned with a large enough dynamic. This can for example be handled with a multiband approach as discussed in this work. In [76] an alternative approach applying compressed sensing for wide-band cognitive radios using sub-Nyquist sampling is investigated.

Spectrum Management

Managing the spectrum assignments in a context where the available spectrum as well as the demand is varying require approaches that combine technical and economical aspects. Especially when the demand for spectrum is greater than the available spectrum mechanisms are needed that consider the value of currently having access to spectrum for each of the participating users or systems. If for example a user has an important and time-critical information to transmit, a transmission opportunity is of higher value than for a user with non-critical data. Due to the similarity to other economical markets also similar approaches can be applied in this case. Especially auctions are a suitable tool spectrum assignments if the demand exceeds the availability. Research in this area covers low-level auctions, that is, very small portions of spectrum are auctioned within a communication system in the MAC layer [53]. But also on a higher level auctions can be applied, for example between different network operators who offer their spare channels temporarily to other operators [31, 86].

2.3.3 Standardization

The IEEE Standards Coordinating Committee 41, Dynamic Spectrum Access Networks, (SCC 41) has amongst other duties the responsibility to vote on approval of proposed IEEE standards and to develop proposed standards in the area of dynamic spectrum access networks [2]. It was originally established by the IEEE Communications Society (ComSoc) and the Electromagnetic Compatibility (EMC) Society as the IEEE 1900 Standards Committee on Next Generation Radio and Spectrum Management in the first quarter of 2005, and was reorganized in the first half of 2007, thereby changing its name.

Currently, within SCC41 the following working groups exist, dealing with different aspects of dynamic spectrum access:

- IEEE 1900.1: Working Group on Terminology and Concepts for Next Generation Radio Systems and Spectrum Management
- IEEE 1900.2: Working Group on Recommended Practice for Interference and Coexistence Analysis
- IEEE 1900.3: Working Group on Recommended Practice for Conformance Evaluation of Software Defined Radio (SDR) Software Modules
- IEEE 1900.4: Working Group on Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks
- IEEE 1900.A: Working Group on Dependability and Evaluation of Regulatory Compliance for Radio Systems with Dynamic Spectrum Access

For further information, the reader is referred to [2].

The currently most advanced standardization effort for communication systems apply dynamic spectrum access is IEEE 802.22 [1]. This standard specifies the PHY and MAC layer for secondary systems operating as wireless regional area networks in frequency that are allocated to TV broadcast services [36]. For the PHY layer OFDM [50] is used, similar to the approach discussed in this work.

3 OFDM based Overlay Systems

On the way to full flexibility regarding spectrum allocation it is not always possible to introduce newly developed dynamic access strategies, as discussed in Chapter 2, right away. There are some scenarios, for example during the transition phase, where it is not possible to avoid blocked spectrum. Thus, the concept of overlay systems is being developed, which allows to exploit the otherwise blocked spectrum by operating an additional and independent system in the same frequency band. To enable a successful coexistence without mutual interference, the overlay system has to work around the licensed system's spectrum occupation. One of the most promising approaches for the design of overlay systems is OFDM [15, 26]. Therefore in this chapter OFDM based overlay systems are introduced and several related issues are discussed.

3.1 System Model

In this work a scenario is assumed where a PU system and a SU system are being operated in the same frequency band. Two assumptions are made for the PU system [51]:

- The PU system has priority and must not be affected by the SU system.
- The PU system must not be modified.

This concludes that all necessary signal processing and coordination resulting from the coexistence must be implemented in the SU system. Therefore, the design of SU systems aims for two main goals [20]:

- Minimize the mutual interference between PU and SU system.
- Utilize the available (idle) resources as good as possible.

To attain these goals the SU system must be able to detect the PU system. As discussed later in more detail, it is assumed that the SU system uses an energy detection approach. Therefore, PU systems based on code division multiple access (CDMA) are not suitable candidates for the discussed overlay scenarios and it is thus required that potential PU systems are using time division multiple access

(TDMA) and frequency division multiple access (FDMA) or a combination of both, which results in detectable spectrum holes. Furthermore, the spectrum access of the SU system must not influence the behavior of the PU system, which is why systems based on carrier sense multiple access with collision avoidance (CSMA/CA) also cannot be used as a PU system due to their carrier sensing property. The channels of the PU system are assumed to be equally-sized and their bandwidth is denoted by Δf_{PU} as depicted in Figure 3.1. The total bandwidth of the frequency band in which one PU is operating is denoted by B_{PU} . A scenario with more than one frequency band and several PU systems is discussed in Chapter 5.

The SU system is assumed to be using OFDM, since with OFDM a highly spectrally-efficient, flexible and sophisticated technology is available. It is already successfully implemented in standards such as IEEE 802.11 [6] and 802.16 [4] and is also shown to be a good candidate for SU systems [36, 84]. One main advantage is that each subcarrier can be switched on and off individually, which allows an easy spectral shaping of the transmit signal, and therefore a dynamic adaptation to the current spectrum access of the PU system. Furthermore, the SU system has to periodically perform spectrum occupation measurements to determine which parts of the spectrum are currently occupied by the PU system. In SU systems using OFDM this can be done efficiently and without much additional cost regarding hardware by using the already existing fast Fourier transform (FFT) component, which is another advantage of this approach. Depending on the exact specification of the SU system the number of subcarriers N_{sc} used for the bandwidth of one PU system channel can vary, and therefore also the subcarrier spacing. The relation between the subcarrier spacing Δf_{SU} and Δf_{PU} is given by $N_{\text{sc}} \cdot \Delta f_{\text{SU}} = \Delta f_{\text{PU}}$. The example shown in Figure 3.1 uses $N_{\text{sc}} = 4$ subcarriers. Note that a smaller subcarrier spacing comes along with a longer symbol duration.

The result of the spectrum occupancy measurement is the spectrum occupancy vector (SOV) with binary elements, indicating for each subcarrier if it is available for the SU system (“1”) or not (“0”). When the SU system uses more than one subcarrier for a PU channel, that is, $N_{\text{sc}} > 1$, the corresponding elements of the SOV can be combined in one element representing the PU channel instead of a subcarrier of the SU system, leading to the compressed SOV (cSOV). This is especially reasonable since the applied detection approach discussed later is based on the joint evaluation of all subcarriers to determine whether the corresponding PU channel is occupied or not. The impact of an inaccurate SOV on the performance of an overlay system based on multi-carrier code division multiple access (MC-CDMA) is discussed in [13].

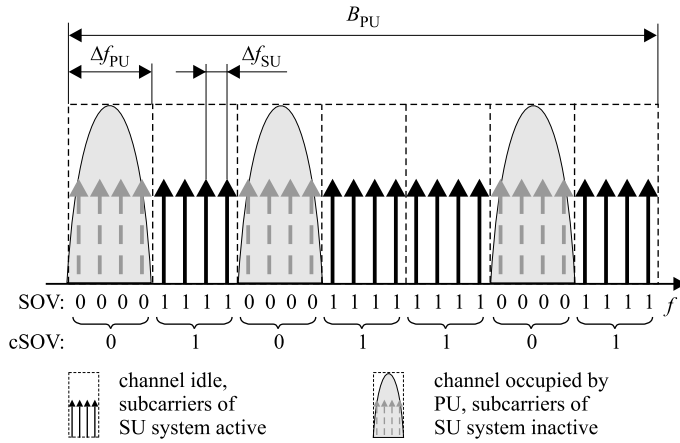


Figure 3.1 Spectrum occupancy vector indicating the idle channels of the PU system and their occupation by the SU system.

3.1.1 Overlay Scenarios

In the following, three possible overlay scenarios in different frequency ranges (very high frequency (VHF) band, 1 GHz and 1.8 GHz) are briefly summarized, serving as illustration. The scenarios were developed within the project Techniken, Algorithmen und Konzepte für COFDM Systeme zur Koexistenz mit autorisierten Systemen im selben Frequenzband (TAKOKO) and are discussed in detail in [16, 25].

Overlay Scenario in the VHF band

The VHF band from 117.975 MHz to 137 MHz is licensed for aeronautical communications and is mainly used for voice communication between aircraft and ground stations. With a channel bandwidth of 25 kHz this results in 760 channels excluding the guard bands, most of which are operated with double sideband amplitude modulation (DSB-AM). For managing the air traffic in a geographical area air traffic control (ATC) sectors are used, to each of which one common VHF channel is assigned. Within the ATC sector communication between controllers and pilots takes place on the channel by following the ‘listen before push-to-talk’ principle, which means that a participant only starts talking when nobody else is talking. This implies that an overlay system based on a detection approach cannot be operated in this frequency band, because the receivers of the DSB-AM system are constantly

listening even if the channel is idle. Therefore, only locally unused channels may be used by the overlay system, which can be derived from the current location of the aircraft. Note that this also excludes active channels in nearby ATC sectors. Nevertheless, estimations of the available bandwidth for the overlay system [16] as well as measurement flights [10] show that only 22 percent of the channels are occupied at the ground of Munich airport, corresponding to an unused bandwidth of about 15 MHz.

This leaves enough bandwidth for an overlay system that can provide an approaching aircraft with up-to-date information regarding the current situation at the airport, including for example an airport map displaying the positions of other aircraft or the approved taxi-way. This overlay system serving as an aeronautical communication and information system is designed as a system similar to a wireless local area network (WLAN) and is based on existing standards [6]. A channel bandwidth of 1 MHz is used and fitting 12 subcarriers in one PU channel results in a subcarrier spacing $\Delta f_{\text{SU}} = 2.0833$ kHz and an FFT period $T_{\text{FFT}} = 480$ μs . The adapted system and OFDM parameters are summarized in Table 3.1. Characteristic for this overlay scenario is that the occupied channels only depend on the location and are only time variant in the eyes of a moving aircraft and not due to a dynamic spectrum access of the PU system. Since in this case the SU system does not need a detection subsystem, the main objectives in this scenario are the mitigation of interference resulting from the PU system as well as the suppression of sidelobes resulting from the SU system's transmitted OFDM signal. Corresponding methods and approaches for reducing the mutual interference have been discussed in detail for example in [28, 29].

Parameter	DSB-AM	DME	GSM
SU channel bandwidth B_{SU}	1 MHz	10 MHz	10.2 MHz
subcarrier spacing Δf_{SU}	2.0833 kHz	10 kHz	40 kHz
FFT length	512	1024	256
occupied subcarriers	108 = 9 · 12	300	varying
available data subcarriers	336	610	varying
FFT period $T_{\text{FFT}} (1/\Delta f_{\text{SU}})$	480 μs	100 μs	25 μs
guard interval T_{GI}	20 μs	10 μs	6.25 μs
total symbol length T_{sym}	500 μs	110 μs	31.25 μs
subcarriers per PU channel N_{sc}	12	100	5

Table 3.1 OFDM parameter configuration for the SU systems in the DSB-AM, DME and GSM scenario.

A flexible communication system coexisting with a DME system

Another class of interesting candidates for PU systems are radar systems. For this scenario the frequency range from 960 MHz to 1215 MHz is considered, which is subdivided into channels with 1 MHz bandwidth used for aeronautical navigation services. The channels are assigned to different services including for example tactical air navigation (TACAN) and secondary surveillance radar (SSR), but the majority is reserved for DME. Disregarding channels used by other services still leaves a large bandwidth potentially available for an SU system. Again estimations show that only about 22 percent of the DME channels are actually occupied when regarding a location on the ground. In DME the airborne transponder basically transmits pulses which are answered by the ground station. Based on the delay of the answer the aircraft can then determine the distance [42]. The main difference to the scenario in the VHF band is that in this case the interference caused by the SU system can be neglected, because the transmit power of the SU system can be assumed to be much smaller than the transmit power of the PU system. Transponders in a DME system have a transmit power up to 2 kW to cover a large range. Furthermore, in this scenario the SU system is required to have a detection subsystem for dynamically detecting the channels used by DME. Note that this usually is only necessary in the initialization phase when deploying the SU system.

The SU system can for example be a flexible communication system for disaster scenarios. Such systems are required to be portable, have to function without infrastructure and must be quickly deployable. The physical layer is adapted to the characteristics of the DME signal and especially the OFDM symbol duration is set to $T_{\text{FFT}} = 100 \mu\text{s}$ so that one radar impulse can destroy a maximum of two symbols. This results in a subcarrier spacing of $\Delta f_{\text{SU}} = 10 \text{ kHz}$ and therefore $N_{\text{sc}} = 100$ subcarriers are used for one DME channel. The large N_{sc} also has the advantage that interference resulting from neighboring DME channels only affects the exterior subcarriers and that interference can be handled individually for each subcarrier for example by adaptive modulation [46] and bit and power loading algorithms [11]. An overview on the OFDM parameters on this scenario is given in Table 3.1.

A MAN system coexisting with a GSM system

In Europe the frequencies from 1710 MHz to 1880 MHz are reserved for mobile communications and are operated with GSM [41], which has a cellular structure with base stations and mobile terminals. Up- and downlink are separated by frequency division duplex (FDD) in two frequency bands of 75 MHz bandwidth per band. Both frequency ranges are subdivided into channels with 200 kHz band-

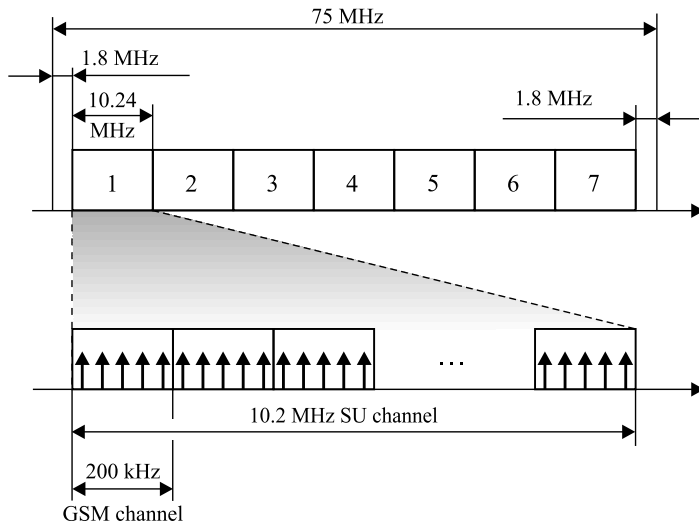


Figure 3.2 Available bandwidth for the MAN based SU system.

width. For the overlay scenario only the downlink is considered, since the base stations are easier to detect due to their fixed position and the constantly active control channels. In order to avoid interference, not all channels are used in each cell and neighboring cells are assigned different channels. Besides the described FDMA component additionally TDMA is used. Each GSM channel is divided into frames containing 8 timeslots with $576.9 \mu\text{s}$. Either Gaussian minimum-shift keying (GMSK) or octal phase shift keying (8-PSK) is used for modulation resulting in a data rate of 270.833 kbit/s . The available bandwidth for an overlay system strongly depends on the specific utilization by the PU system's customers, which is also dependent on the time of day. The interesting challenge in this scenario is the fast detection of spectrum holes. Due to the combination of TDMA and FDMA the spectrum holes can be very small, in the worst case they have the size of one timeslot.

In order to utilize the gaps within the PU system, in this scenario a fixed wireless radio system, for example a metropolitan area network (MAN), is used. These systems are for example used to provide wireless data connections to house, substituting the digital subscriber line (DSL) and thus avoiding expensive rental costs for the "last mile". Usually fixed outdoor antennas are used, whereas in-house conventional technologies like Ethernet/WLAN are applied to distribute the data, resulting

in a transparent connection for the user. A GSM provider can either use the overlay system for own additional services or sell the unused capacity to another provider. In the first case it might be possible to use the same infrastructure, whereas in the latter case an independent provider also requires an independent overlay system. There are already several standards for MANs, so it is reasonable to stay as close as possible to an existing standard and only modify necessary parameters for operation in an overlay scenario. Nevertheless, it is obvious that due to the changing allocation the resulting system still will not be compatible to the original standard. For this overlay system the physical layer parameters are derived from IEEE 802.16 WirelessMAN-OFDM, which is an OFDM based option of 802.16 [4]. It uses 256 subcarriers, and for modulation quadrature phase shift keying (QPSK), 16-quadrature amplitude modulation (QAM) or 64-QAM is allowed. For duplex either FDD or time division duplex (TDD) can be applied.

The design of the physical layer parameters are constrained by the underlying standard as well as the coexistence environment:

- The shorter the OFDM symbols are, the shorter gaps in time domain can be utilized.
- An integer number of subcarriers per GSM channel has to be used.
- The number and bandwidth of the MAN channels should be optimized regarding the totally available bandwidth, which has to be divided into 2^n , $n \in \mathbb{N}$ MAN channels with at least 1.25 MHz bandwidth.

A convenient number of OFDM subcarriers per GSM channel is for example $N_{sc} = 5$ as depicted in Figure 3.2. This results in a subcarrier spacing $\Delta f_{SU} = 40$ kHz with a corresponding FFT-length $T_{FFT} = 25 \mu s$. Using a guard interval of $T_{GI} = 6.25 \mu s$ results in a total symbol length of $T_{sym} = 31.25 \mu s$. When applying a 256-FFT, this leads to a channel bandwidth in the SU system of $B_{SU} = 10.2$ MHz, and therefore to 7 possible channels within the total bandwidth of the GSM downlink. Thus, the channel can be efficiently addressed with 3 bits. This scenario certainly is not suitable for fully loaded PU systems as they for example exist in cities. But in rural areas where more capacity is available for the PU system and where connecting houses via cable is expensive, this SU system can save costs and increase spectral efficiency. Note that the transmission range of an SU system has to be small compared to the transmission range of the PU system. In rural areas GSM cells can have a radius of up to 35 km.

3.1.2 Characteristic Parameters of Primary User Systems

In order to design an optimal SU system, it is necessary to know the behavior of the PU system. In this section an overview on some important characteristic parameters is given that influence and constrain the SU system.

PU systems coming under consideration for an overlay scenario strongly differ as illustrated in the previous section. Besides the used frequency bands these systems especially vary in the access behavior resulting in different types of spectrum holes. The GSM system uses a channel bandwidth of 200 kHz and timeslots of $577 \mu\text{s}$ duration, whereas the DSB-AM system utilizes 25 kHz channels. The channel allocation is constant at a specific airport and only changes relatively to the SU system while the aircraft is moving, resulting in very slow changes. Regarding the variety of these different systems, it is interesting to find a set of characteristic parameters with which it is possible to generally describe important aspects of PU systems in the context of an overlay environment and to derive guidelines for designing OFDM based overlay systems.

The basic resource for an overlay system is the unused spectrum in a defined frequency band. Since each transmission requires a bandwidth as well as a certain time, it is necessary to regard the time/frequency plane. The spectrum utilization of the PU system is denoted by η_{PU} and is defined as the fraction of the available resources (B_{PU}) that are used in the regarded period of time. The remaining resources (that is, the spectrum holes) are the capacity of the SU system

$$C_{\text{SU}} = 1 - \eta_{\text{PU}}. \quad (3.1)$$

Depending on the PU system's access the unused spectrum appears to be fragmented resulting in spectrum holes of different sizes and shapes. In the general case the width of the spectrum holes can vary and must be estimated. Nevertheless, here it is assumed that the PU system uses an equidistant channel spacing Δf_{PU} . Thus, the unused spectrum can be described by blocks in the time/frequency plane of equal width Δf_{PU} and varying length as depicted in Figure 3.3. The PU system's access is assumed to be a random process, so that in consequence the length of the spectrum holes is also random. It is described by the random variable τ with the expectation $E\{\tau\}$ and the variance $D^2\{\tau\}$. In some cases, for example in a system based on TDMA, τ has a minimum value τ_{min} which is equal to the used TDMA slot length.

Besides the length of the spectrum holes it is also important how often the PU system changes its spectrum access and thus how many spectrum holes occur in the

Symbol	Parameter
B_{PU}	PU system's bandwidth
Δf_{PU}	channel spacing
η_{PU}	PU system's spectrum utilization
C_{SU}	resulting SU system's spectrum capacity
R_{PU}	access rate
$E\{\tau\}$	expected length of spectrum holes
τ_{min}	minimum length of spectrum holes

Table 3.2 Characteristic parameters of PU systems.

regarded time. Even if the capacity of the SU system remains constant, it is a difference for the detection subsystem whether the PU system produces many small holes or only a few but large spectrum holes, resulting in a larger or smaller degree of fragmentation, respectively. A measure for the fragmentation is the access rate R_{PU} of the PU system. R_{PU} gives the number of beginning transmissions of the PU system in the complete bandwidth B_{PU} per second. Hereby, two transmissions in different channels beginning at the same time (as for example possible in TDMA based systems) are considered as a single event. This has the advantage that spectrum holes spanning more than one PU channel do not increase the access rate, since they are also not contributing to the fragmentation of the spectrum. Note that $E\{\tau\} < \frac{N_{\text{PU}}}{R_{\text{PU}}}$ with N_{PU} denoting the number of PU channels if the access is distributed uniformly over all PU channels.

In a scenario where the transmit power of the SUs is high enough to interfere with the PU system, a further parameter is vital. The maximum interference time $T_{\text{IF}_{\text{max}}}$ is the time within which the SU system must release the spectrum and make it available for the PU system. It represents the maximum time that the PU can tolerate interference. Table 3.2 finally gives an overview of the characteristic parameters of PU systems introduced in this section.

3.1.3 Design Parameters of OFDM Secondary User Systems

Designing an overlay system can be compared to laying an optimal grid in the time/frequency plane over the PU system as shown in Figure 3.3. This equals a discretization of the spectrum in both, the time domain and the frequency domain. Since the discretization steps in time and frequency domain are not independent, the SU system's parameters cannot be chosen arbitrarily. The time reserved for

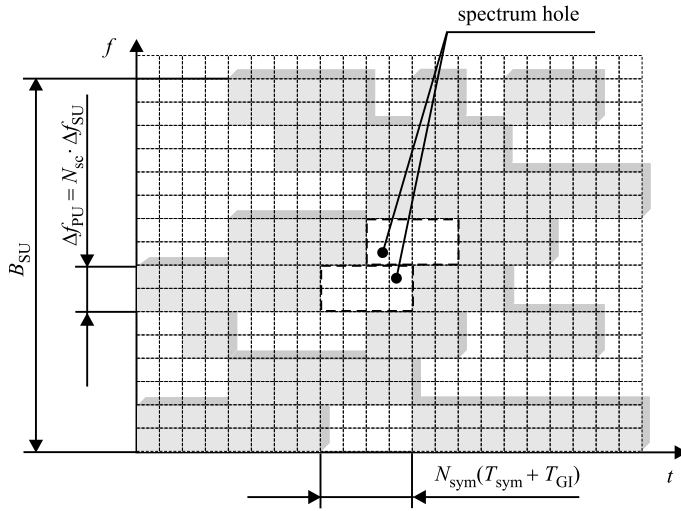


Figure 3.3 Time and frequency constraints for the configuration of an OFDM overlay system.

a spectrum occupation measurement including signaling is called detection period T_{dp} and the interval between two detection periods is referred to as update interval T_{up} . Note that the detection period can contain different phases, for example the actual detection phase, the boosting phase or the broadcasting phase, as explained in Section 4.3.1. The update interval is bounded by the maximum amount of time for which a PU system can tolerate interference when occupying new parts of the spectrum, denoted as $T_{IF_{max}}$. Thus, the SU system has to detect a spectrum access of the PU system and stop transmitting within $T_{IF_{max}}$ resulting in the time domain constraint:

$$T_{up} \leq T_{IF_{max}}. \quad (3.2)$$

This leads to the following constraint regarding the number of OFDM symbols fitting in one update interval with T_{sym} denoting the duration of one OFDM symbol and T_{GI} the guard interval, respectively:

$$N_{sym} \cdot (T_{sym} + T_{GI}) \leq T_{IF_{max}}. \quad (3.3)$$

In frequency domain this results in a minimum subcarrier spacing of

$$\Delta f_{SU} \geq \frac{N_{sym}}{T_{IF_{max}} - N_{sym}T_{GI}}. \quad (3.4)$$

Symbol	Parameter
Δf_{SU}	subcarrier spacing
T_{S}	symbol length
N_{FFT}	FFT-size
T_{GI}	guard interval length
B_{SU}	total bandwidth
N_{sc}	subcarriers per channel
T_{up}	update interval
N_{sym}	number of symbols per update interval

Table 3.3 Design parameters for OFDM overlay systems.

On the other hand the SU system's subcarrier spacing must not be larger than the PU system's channel bandwidth, resulting in the frequency domain constraint:

$$\Delta f_{\text{SU}} \leq \Delta f_{\text{PU}}. \quad (3.5)$$

Furthermore, when using more than one subcarrier for a PU channel, the used number of subcarriers N_{sc} must be a natural number resulting in the following possible values for the subcarrier spacing:

$$\Delta f_{\text{SU}} = \frac{\Delta f_{\text{PU}}}{N_{\text{sc}}}, \quad N_{\text{sc}} \in \mathbb{N}. \quad (3.6)$$

Based on these two constraints for the subcarrier spacing and with the totally available overlay bandwidth B_{SU} it is possible to derive the required FFT-size:

$$N_{\text{FFT}} = 2^n \geq \frac{B_{\text{SU}}}{\Delta f_{\text{SU}}}, \quad n \in \mathbb{N}. \quad (3.7)$$

The OFDM parameters for SU systems are summarized in Table 3.3.

3.2 Detection Subsystem

For the coexistence of a SU system and a PU system in the same frequency band, reduction of mutual interference plays a vital role. This is the objective of the detection subsystem, which therefore is a central component of an SU. The spectrum occupation of the PU system in the time/frequency plane—as depicted in Figure 3.3—shows that the spectrum holes are limited in time as well as in frequency

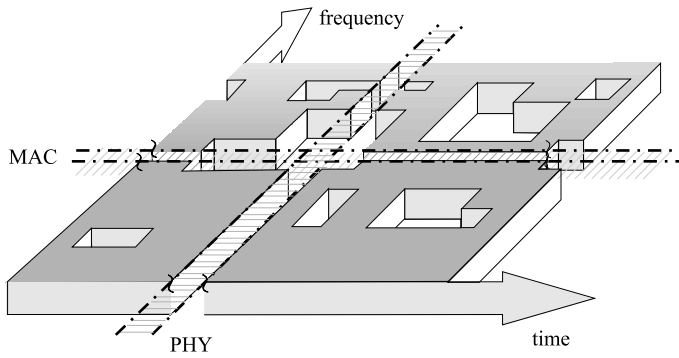


Figure 3.4 Spectrum occupation of the PU system in the time-frequency plane and the directions where the PHY/MAC layers protect the PU system from interference.

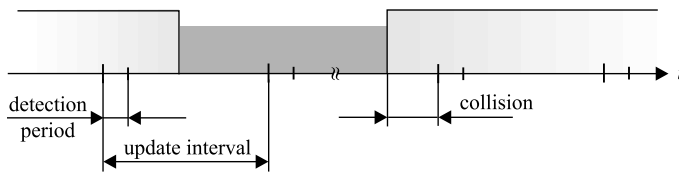


Figure 3.5 Detection period and update interval.

direction. In order to successfully exploit the spectrum holes mutual interference in both dimensions has to be considered. In this context, physical and MAC layer deal with different aspects of interference reduction.

The physical layer is responsible for interference mitigation in frequency direction, that is, interference resulting from out-of-band radiation between active subchannels of the PU system and adjacent subcarriers of the SU system. This can be illustrated by keeping time in Figure 3.4 constant and performing a cut parallel to the frequency axis. The resulting figure is the already discussed Figure 3.1 in Section 3.1. Interference resulting from the SU system can for example be handled with sidelobe suppression techniques [27].

In contrast, the MAC layer is responsible for the avoidance of interference in time direction. This is visualized when looking again at Figure 3.4. Now keeping the frequency constant—that is, taking a snapshot of a single channel of the PU system—and performing a cut parallel to the time axis, Figure 3.5 is derived. Interference

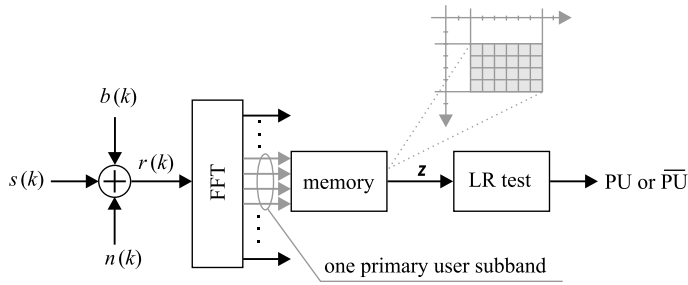


Figure 3.6 The detector model.

occurs in form of collisions when both, the PU system and the SU system, are transmitting at the same time. The SU system has to reduce the interference to a minimum, so that it can still be handled by the PU system. This is done by performing the spectrum occupancy measurements in intervals that are small enough to limit the time for possible collisions. The MAC layer is responsible for initiating the spectrum occupation measurements, and thus also for determining the update interval. This directly influences the possible interference as well as the performance of the SU system. A framework for optimizing both layers in a cross-layer context is proposed in [14].

In the following, the detector model used in the remainder of this work is introduced.

3.2.1 The Detector Model

Figure 3.6 shows the detector model [49] which is used for detecting the PU system at each SU as well as for receiving the signaling results in the collection phase (see Section 4.3.1) at the secondary user access point (SU-AP), with the difference, that in the latter case additionally the boosting signal is present. The received signal $r(k)$ is composed of the PU signal $s(k)$ and the additive white Gaussian noise (AWGN) component $n(k)$:

$$r(k) = s(k) + n(k) \quad (3.8)$$

with the prerequisite that in the detection phase all SUs are silent so that only the PU signal is detected. For the sake of simplicity, now only a single channel of the PU system is considered. Assuming a no line of sight (NLOS) scenario between the transmitting PU and the detecting SU, the received signal $s(k)$ is composed of

many multipath signals and can therefore be considered as a zero-mean complex Gaussian random variable when applying the central limit theorem. Furthermore, the boosting signal $b(k)$ is assumed to be zero-mean Gaussian distributed, so that finally the total received signal $r(k)$ also is a complex zero-mean Gaussian random variable. The signal $r(k)$ is then transformed into the frequency domain with an FFT, yielding the vector

$$\mathbf{z} = (x_1, x_2, \dots, x_M; y_1, y_2, \dots, y_M)^T \quad (3.9)$$

with $2M$ elements representing the $2M$ samples resulting from one detection phase stored in the detector's memory. Note that M depends on the number of subcarriers N_{sc} used for a PU channel as well as the number of detection cycles within a detection phase. Due to the linear property of the FFT the elements of \mathbf{z} are zero-mean Gaussian random variables and are represented by the complex variables

$$z_m = x_m + jy_m = R(m) = S(m) + N(m); m = 1, 2, \dots, M \quad (3.10)$$

where x_m and y_m denote the real and imaginary parts of the m^{th} component of \mathbf{z} . In (3.10), $\{S(m) = S_x(m) + jS_y(m)\}_{m=1}^M$ and $\{N(m) = N_x(m) + jN_y(m)\}_{m=1}^M$ represent the signal components in frequency domain resulting from the PU system and noise, respectively. The mean powers of $S(m)$ and $N(m)$ are denoted by $2\sigma_S^2$ and $2\sigma_N^2$.

To derive the detection probability and the false alarm probability for the detection of a PU system, defined as

$$\begin{aligned} P_{\text{dct}} &= \int_{\lambda_1}^{\infty} f_{\Lambda|\text{PU}}(\lambda|\text{PU})d\lambda, \\ P_{\text{false}} &= \int_{\lambda_1}^{\infty} f_{\Lambda|\overline{\text{PU}}}(\lambda|\overline{\text{PU}})d\lambda, \end{aligned} \quad (3.11)$$

first the conditional probability density functions (pdfs) $f_{\Lambda|\text{PU}}(\lambda|\text{PU})$ and $f_{\Lambda|\overline{\text{PU}}}(\lambda|\overline{\text{PU}})$ are derived, with PU denoting the event that a PU is detected and $\overline{\text{PU}}$ the negated event. Then the Neyman-Pearson strategy [78] is applied, resulting in a likelihood ratio (LR) test with the threshold λ_1 . The detailed steps for these calculations are omitted here, but can be found in [81]. The exact LR test depends on the signal statistics and whether the detector matches them or not. In this work all $S(m)$ and $N(m); m = 1, 2, \dots, M$ are assumed to be independent and the

detector to be matched, resulting in the LR test

$$\Lambda(\mathbf{z}) = \mathbf{z}^T \mathbf{z} = \sum_{m=1}^M x_m^2 + \sum_{m=1}^M y_m^2 \stackrel{\text{PU}}{\underset{\text{P}\bar{\text{U}}}{>}} \lambda_1. \quad (3.12)$$

This corresponds to the best case scenario which is motivated by the theoretical bounds studied in Section 4.3.2. A detailed discussion of the best and worst cases, that is, a matched and unmatched detector, for correlated and uncorrelated signals can be found in [81]. Since $\Lambda(\mathbf{z})$ is a sum of $2M$ squared identically distributed Gaussian random variables, it is centrally χ^2 -distributed [78] and the corresponding conditional pdfs turn out as

$$f_{\Lambda|\text{PU}}(\lambda|\text{PU}) = \begin{cases} \frac{\lambda^{M-1}}{2^M(\sigma_S^2 + \sigma_N^2)^M(M-1)!} \cdot \exp\left(-\frac{\lambda}{2(\sigma_S^2 + \sigma_N^2)}\right) & \text{for } \lambda \geq 0 \\ 0 & \text{for } \lambda < 0 \end{cases} \quad (3.13)$$

and $f_{\Lambda|\text{P}\bar{\text{U}}}(\lambda|\text{P}\bar{\text{U}})$, which can be derived from (3.13) by setting $\sigma_S^2 = 0$. By integration and with the help of the integral table [44] this finally results in the single user detection probability

$$P_{\text{dct}} = \frac{1}{2^M(\sigma_S^2 + \sigma_N^2)^M(M-1)!} \cdot \exp\left(-\frac{\lambda_1}{2(\sigma_S^2 + \sigma_N^2)}\right) \cdot \left(2(\sigma_S^2 + \sigma_N^2)\lambda_1^{M-1} + \sum_{m=1}^{M-1} (2(\sigma_S^2 + \sigma_N^2))^{m-1} \lambda_1^{M-m-1} \prod_{l=1}^m (M-l)\right) \quad (3.14)$$

and the single user false alarm probability ($\sigma_S^2 = 0$ in (3.14))

$$P_{\text{false}} = \frac{1}{(2\sigma_N^2)^M(M-1)!} \exp\left(-\frac{\lambda_1}{2\sigma_N^2}\right) \cdot \left(2\sigma_N^2\lambda_1^{M-1} + \sum_{m=1}^{M-1} (2\sigma_N^2)^{m-1} \lambda_1^{M-m-1} \prod_{l=1}^m (M-l)\right). \quad (3.15)$$

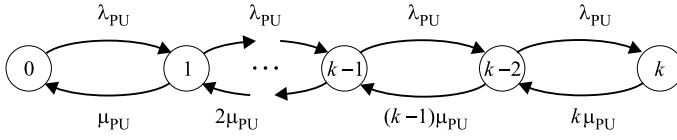


Figure 3.7 Markov chain modeling the number of occupied channels in a PU system.

3.3 Traffic Model for Primary User Systems

In the following, a simple traffic model for the PU system based on queuing theory is discussed. Based on this simple model, the optimal length of the update interval in the SU system is investigated in Section 3.4. Furthermore, it is used in Section 3.5.2 to derive the available bandwidth of an SU system with SUs in two neighboring cells of the PU system, each with a different traffic load. In the second step, the simple traffic model is extended regarding a PU system based on TDMA similar to GSM. Incoming calls are mapped to different time slots, resulting in a different and much more fragmented spectrum occupation.

3.3.1 Simple Model

The spectrum occupation of one PU system is modeled as an $M/M/k/k$ loss system. Incoming traffic arrives according to a Poisson process with the total arrival rate λ_{PU} . The inter arrival times are exponentially distributed. Service times also follow an exponential distribution with parameter μ_{PU} which is the service rate corresponding to a mean call or packet duration of $\frac{1}{\mu_{PU}}$. This model is especially used for voice traffic, but also can be applied for packet based traffic when considering the aggregate traffic of a large number of similar users [22]. The PU system is assumed to have k equal channels and incoming traffic is mapped randomly to the available channels. If there is no free channel the current traffic request is dropped without reattempting a transmission later on, that is, there is no waiting queue. With these assumptions the spectrum access of the PU system can be modeled as a Markov chain as shown in Figure 3.7. The steady state probabilities are denoted by

$$\boldsymbol{\pi} = [\pi_0, \pi_1, \dots, \pi_k] \quad (3.16)$$

and indicate the probabilities for n channels being simultaneously occupied, with $n = 1, 2, \dots, k$. They can be derived from the global balance equation

$$\lambda_{PU} \cdot \pi_{n-1} = n \cdot \mu_{PU} \cdot \pi_n, \quad (3.17)$$

and result in

$$\pi_n = \pi_0 \left(\frac{\lambda_{\text{PU}}}{\mu_{\text{PU}}} \right)^n \frac{1}{n!}, \quad n = 1, 2, \dots, k. \quad (3.18)$$

Since the sum of all probabilities has to fulfill the condition $\sum_{n=0}^k \pi_n = 1$, π_0 is given by

$$\pi_0 = \left[\sum_{n=0}^k \left(\frac{\lambda_{\text{PU}}}{\mu_{\text{PU}}} \right)^n \frac{1}{n!} \right]^{-1}. \quad (3.19)$$

Note that the blocking probability is given by π_k , which is also known as Erlang B formula when inserting (3.19) in (3.18).

Finally, the probability that l channels are available for the SU system can be easily derived from (3.18):

$$\bar{\pi}_l = \pi_{k-l}. \quad (3.20)$$

3.3.2 Model for a TDMA Based Primary User System

Now a GSM-like PU system is considered which is used for speech transmissions and is based on TDMA. Similar to the previous section, the PU system contains k channels, but now they are divided into frames of length T_F with N_{slot} slots. The complete PU system can handle up to $k \cdot N_{\text{slot}}$ calls in parallel and the total number of incoming voice calls is distributed equally among the available channels and time slots. Hence, the arrival rate at each channel is

$$\lambda_{\text{CH}} = \frac{\lambda_{\text{PU}}}{k}. \quad (3.21)$$

It is further assumed that the frame size is small compared to the duration of a call, that is, $\frac{1}{\mu_{\text{PU}}} \gg T_F$. In order to determine the unused spectrum, first a single channel is considered. The model will then be extended to multiple channels [19].

Single Channel ($k = 1$)

A channel with N_{slot} time slots per frame can also be modeled as a $M/M/N_{\text{slot}}/N_{\text{slot}}$ queuing system, since it can handle up to N_{slot} calls simultaneously. Therefore also the steady state probabilities—denoted by $\hat{\pi} = [\hat{\pi}_0, \hat{\pi}_1, \dots, \hat{\pi}_{N_{\text{slot}}}]$ —are similar and can be derived from (3.18) by substituting λ_{PU} with λ_{CH} . The probability $\hat{\pi}_1$ for

instance gives the probability that exact one call is taking place and thus only one time slot is occupied in each frame. Provided that the used time slot is chosen randomly in every frame, the probability for a specific time slot being occupied is given by

$$P_{\text{occ}} = \sum_{i=0}^{N_{\text{slot}}} \frac{i}{N_{\text{slot}}} \hat{\pi}_i, \quad (3.22)$$

and

$$P_{\text{occ}} = \sum_{i=0}^{N_{\text{slot}}} \frac{N_{\text{slot}} - i}{N_{\text{slot}}} \hat{\pi}_i, \quad (3.23)$$

for an available time slot. With equations (3.18) and (3.19) this results in

$$P_{\text{occ}} = \frac{\sum_{i=0}^{N_{\text{slot}}} \frac{N_{\text{slot}} - i}{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}}{\sum_{i=0}^m \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}}. \quad (3.24)$$

Equation (3.24) gives the probability that a specific timeslot is available for an overlay transmission.

Multiple Channels ($k > 1$)

Considering multiple channels, the number of simultaneously available time slots N_{free} is given by a binomial distribution

$$P(N_{\text{free}} = l) = \binom{k}{l} P_{\text{occ}}^l (1 - P_{\text{occ}})^{k-l} \quad (3.25)$$

with the expectation

$$E\{N_{\text{free}}\} = k \cdot P_{\text{occ}} = k \cdot \frac{\sum_{i=0}^{N_{\text{slot}}} \frac{N_{\text{slot}} - i}{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}}{\sum_{i=0}^{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}} \quad (3.26)$$

and the variance

$$\begin{aligned} D^2\{N_{\text{free}}\} &= k \cdot P_{\text{occ}}(1 - P_{\text{occ}}) \quad (3.27) \\ &= k \cdot \frac{\sum_{i=0}^{N_{\text{slot}}} \frac{N_{\text{slot}} - i}{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}}{\sum_{i=0}^{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}} \cdot \left(1 - \frac{\sum_{i=0}^{N_{\text{slot}}} \frac{N_{\text{slot}} - i}{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}}{\sum_{i=0}^{N_{\text{slot}}} \left(\frac{\lambda_{\text{CH}}}{\mu_{\text{PU}}} \right)^i \frac{1}{i!}} \right). \end{aligned}$$

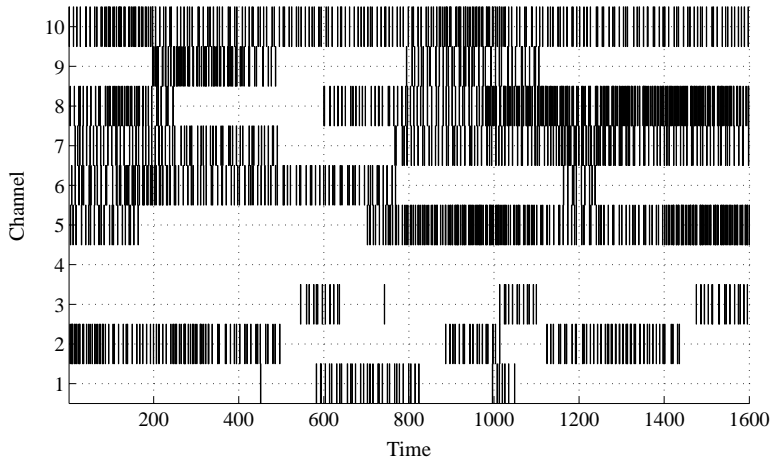


Figure 3.8 Spectrum access of a PU system with $k = 10$, $N_{\text{slot}} = 8$, $\mu_{\text{PU}} = 0.025$, $\lambda_{\text{CH}} = 0.0225$.

Following the Moivre-Laplace's limit theorem this distribution converges towards a normal distribution for large k . An example for the spectrum access of a PU system with $k = 10$ channels and $N_{\text{slot}} = 8$ time slots is shown in Figure 3.8.

3.4 Optimizing the Secondary User System's Parameter Configuration

With the introduced model for the detector and the model for the PU system's spectrum access, it is now possible to perform some optimizations regarding the configuration of the SU system's detection subsystem. The objective of the detection subsystem is to find occurring spectrum holes as accurate as possible and provide information for the other layers when and where the SU can transmit data resulting in a minimum of interference and collisions. In other words, the detection subsystem is responsible for exploiting the capacity C_{SU} of the SU system as good as possible.

There are mainly two parameters that can be adapted to optimize detection. First, the time within an update interval that is used for detection can be optimized. That

is, a trade-off has to be made between spending time for a reliable detection of spectrum holes, and thus reducing the false alarm probability, and exploiting the available time for transmitting data. The second parameter that can be optimized—depending on the PU system's spectrum access—is the length of the update interval. From the SU system's point of view a large update interval reduces the time spent for detection, but also increases the probability that resources are not detected when the PU system releases spectrum shortly after the detection period of the SU system.

3.4.1 Duration of the Detection Phase

Recall from Section 3.2 that the detection process is performed by a cluster of FFT operations, resulting in M complex valued samples in the memory of the detector. This is illustrated in Figure 3.9. The smallest unit in time domain is the length of one OFDM symbol. Each slot can be either used to perform a detection or to transmit/receive data (neglecting the guard interval). The performance of the detector depends on M , that is, the more samples are available, the smaller the false alarm probability is, assuming a given detection probability. However, a larger M requires more time for the detection phase, and therefore also reduces the number of OFDM symbols that are available for transmitting data within each update interval. Therefore it is important to find a good tradeoff between spending time for detection—thus decreasing the false alarm probability—and using the available time for transmitting data.

The efficiency of the detection subsystem regarding one update interval is defined as the fraction of the available resources in an update interval that are effectively expected to be available also for transmission

$$\eta_{DS} = (1 - P_{\text{false}}(N_D \cdot N_{\text{sc}})) \cdot \left(1 - \frac{N_D}{N_{\text{sym}}}\right), \quad (3.28)$$

that is, excluding time spent for spectrum occupancy measurements and false alarms of the detector. N_D is the number of symbols used for detection and N_{sym} the total number of symbols in an update interval as introduced in Section 3.1.3. For the sake of simplicity only a single PU channel is regarded and it is assumed that only one subcarrier is used for a PU channel ($N_{\text{sc}} = 1$ and thus $\Delta f_{\text{SU}} = \Delta f_{\text{PU}}$). Figure 3.10 and Figure 3.11 show the resulting efficiency of the detection subsystem depending on the number of symbols used for the detection for different lengths of the update interval and for different signal-to-noise ratios (SNRs), respectively. For a small number of symbols in the detection phase the influence of the detector model dominates and only adding a few samples to the detection phase already

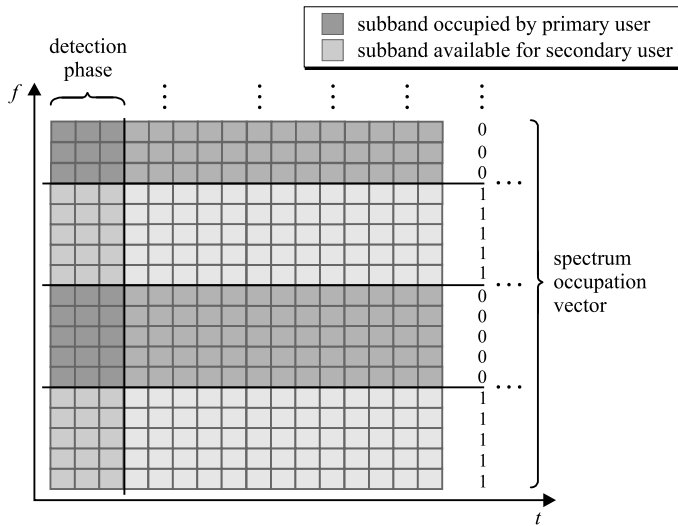


Figure 3.9 Spectrum occupation of a PU in the time-frequency plane with the resulting SOV and three symbols used for the detection phase.

significantly increases the efficiency. This only works up to a certain point, representing the maximum achievable efficiency, where the curves start approximating a straight line with a negative gradient. Now the false alarm probability is very small but the efficiency decreases because an increasing number of symbols is used for the detection without relevantly decreasing the false alarm probability. The gradient of the straight line depends on N_{sym} , but the SNR has a greater effect on the position of the maximum than N_{sym} . This is also illustrated in Figure 3.12. The optimal number of symbols used for the detection phase is depicted depending on the total number of symbols in each update interval and for different SNRs. As expected, for high SNRs significantly less symbols must be used for the detection phase to achieve the optimum efficiency of the detection subsystem. From these results it can be concluded that it is reasonable to consider a dynamic allocation of the symbols to the detection phase when designing the overlay system if the SNR is dynamic. This should also be done when the update interval is changing, although the optima for different update intervals are located close together in comparison. Nevertheless, the achievable efficiency significantly increases for larger update intervals as well as for better SNRs.

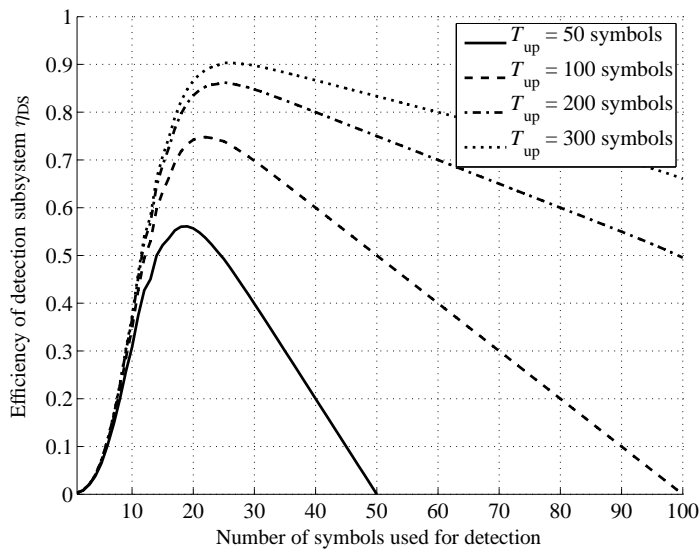


Figure 3.10 Efficiency of the detection subsystem for several update intervals depending on the number of symbols used for detection (SNR = 3 dB).

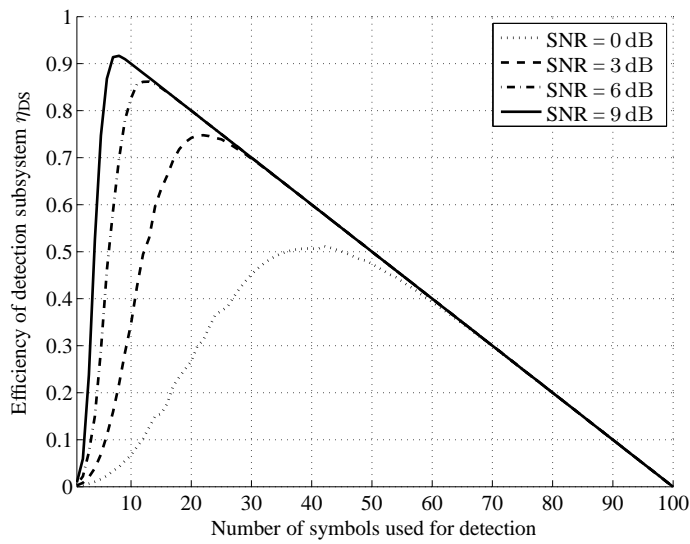


Figure 3.11 Efficiency of the detection subsystem depending on the number of symbols used for detection for different SNR and $N_{sym} = 100$ symbols.

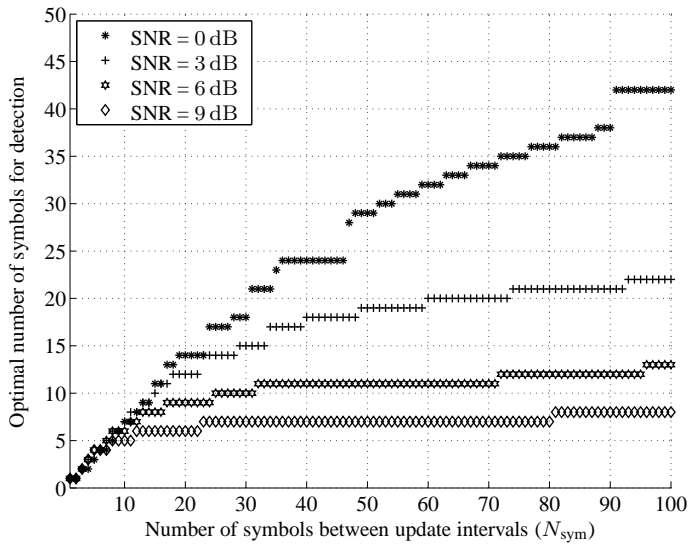


Figure 3.12 Number of symbols resulting in a maximized efficiency of the detection subsystem depending on the total number of symbols within an update interval and for different SNRs.

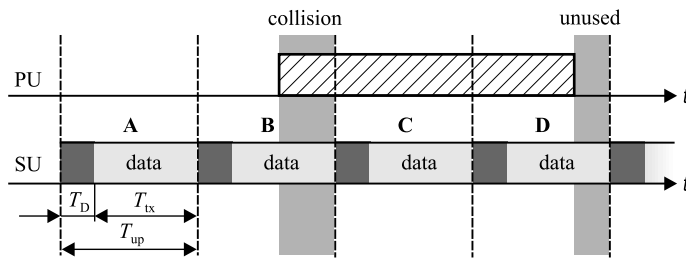


Figure 3.13 A PU system allocating one channel and a SU system detecting it, showing the four basic situations that can occur.

3.4.2 Update Interval

From the perspective of the SU system a long update interval reduces the time spent for detection (and thus more time is available for the actual transmissions), but also increases the probability for collisions with the PU system (there leading to more destroyed data blocks).

Since the SU system only observes the PU system's spectrum occupation during the detection phase, it does not notice any possible change in the PU system's spectrum occupation until the next detection phase is performed. Figure 3.13 displays the four different resulting situations that can occur [20]:

- **A:** The SU system does not detect a spectrum access and the PU system does not occupy the channel during the following time. This is the best situation for the SU system because its assumption regarding the subchannel state is valid for the whole time.
- **B:** The SU system does not detect a spectrum access, but before the next detection phase the PU system occupies the subchannel. This results in a collision between the PU and SU system and part of the SU system's data is destroyed.
- **C:** This situation is similar to A, but now the subchannel is not available to the SU system.
- **D:** The SU system detects a PU spectrum access, but the PU system releases the channel before the next detection phase. This does not result in a collision or destruction of data, but the actually available subchannel is neither used by the PU nor the SU system, degrading the efficiency in spectrum use.

So from the overlay system's point of view a shorter update interval reduces the probability of collisions with the PU system, and therefore also interference, as well as the waste of actually available resources. A longer update interval, however, reduces the time spent for detection phases as well as the signaling effort, since each detection phase goes along with the distribution of the current spectrum occupation vector within the SU system.

Collisions resulting from B-situations can be handled by retransmitting the destroyed packets and only the information transmitted twice during unoccupied times contributes to a performance degradation in the SU system. A more significant impact results from D-situations where available spectrum is not detected. The influence of the update interval length on the efficiency and the potential of its optimization is illustrated by the simulations shown in Figure 3.14. Using the simple traffic model introduced in Section 3.3.1 the efficiency of the detection subsystem is depicted for different arrival rates λ_{PU} depending on the length of the update interval. A fixed detection time $T_{\text{D}} = 0.01$ s including signaling is assumed and the service rate is $\mu_{\text{PU}} = 0.1$. Furthermore, blocked calls are lost. In contrast to the previous section here perfect detection is assumed and the efficiency is not

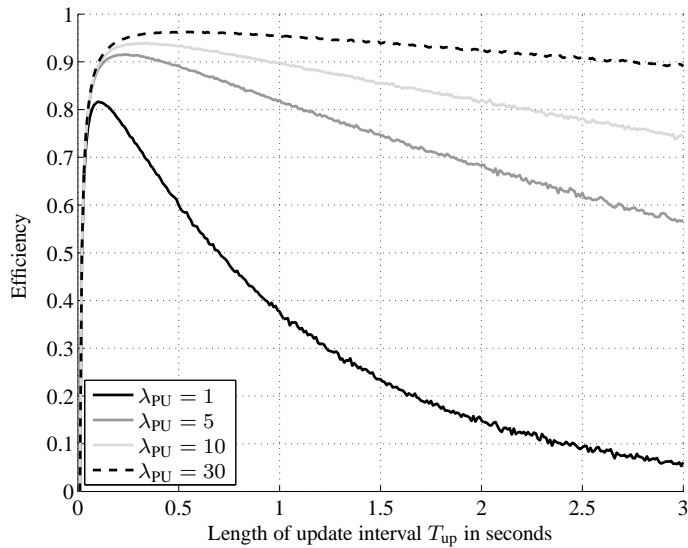


Figure 3.14 Efficiency of the detection subsystem depending on the PU system's traffic.

only considered regarding a single update interval but especially includes the D-situations. For update intervals that have a similar length as the detection period efficiency is very small, since the system spends most of the time for detection as already investigated in the previous section. Therefore, the efficiency increases rapidly with a growing update interval until it reaches a climax. Then it slowly decreases again. This effect results from the D-situations where unused spectrum is not detected.

3.5 Multiple Spectrum Occupation Zones

An area with the same spectrum occupation due to the PU system's access is called spectrum occupation zone. Usually this area depends on the coverage area of the PU system, for example when regarding the downlink of a GSM system as discussed in Section 3.1.1 or a frequency range used for broadcasting television signals. Nevertheless, it is not equal to the coverage area which is defined as the area in which a PU receiver still can receive signals from the PU transmitter (why all PU receivers can be assumed to be located within the coverage area). In contrast, the

occupation zone is defined as the area in which a SU can detect the PU signal with energy detection. The spectrum occupation zone therefore includes the coverage area, but is larger.

In this section first the maximum allowable transmission range for SUs is derived. In the second part two adjacent spectrum occupation zones are considered and the amount of available resources of an SU system with stations in both zones is discussed.

3.5.1 Maximum Allowable Transmission Range of Secondary Users

Figure 3.15 shows a PU system base station. The coverage area and the spectrum occupation zone are assumed to be discs with the base station in the origin. They are denoted by the dashed line and the continuous line, respectively. Furthermore an SU with the corresponding coverage area and spectrum occupation zone is depicted. The radii of the coverage areas are denoted by r_{PU} and r_{SU} for the PU system and the SU system, respectively. With Δ_{PU} and Δ_{SU} denoting the additional fraction of r_{PU} and r_{SU} , the radii of the spectrum occupation zones are given by

$$r_{\text{PU}}^* = (1 + \Delta_{\text{PU}}) \cdot r_{\text{PU}} \quad (3.29)$$

and

$$r_{\text{SU}}^* = (1 + \Delta_{\text{SU}}) \cdot r_{\text{SU}}, \quad (3.30)$$

respectively. Combining (3.29) and (3.30) yields the condition

$$r_{\text{SU}} \leq \frac{\Delta_{\text{PU}} r_{\text{PU}}}{1 + \Delta_{\text{SU}}} \quad (3.31)$$

for the maximal allowable transmission range of the SU. This is also illustrated by the scenario depicted in Figure 3.15 which represents the extreme case. The SU is located on the fringe of the PU system's spectrum occupation zone and thus still can detect the PU system's current spectrum occupation. If it moves a little outside of the occupation zone, it cannot detect the PU system anymore. Nevertheless, the spectrum occupation zone of the SU must not intersect with the PU system's coverage area, yielding the derived maximum for the transmission range.

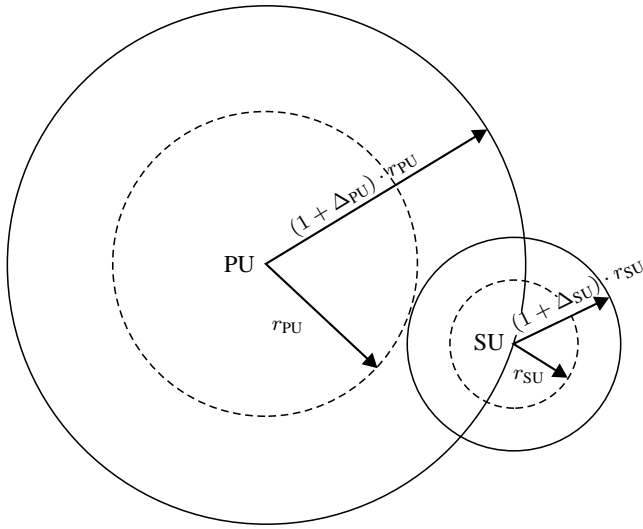


Figure 3.15 Occupation zones and coverage areas of a PU and SU.

3.5.2 Secondary User Systems Operating in Multiple Spectrum Occupation Zones

Now two neighboring spectrum occupation zones with a different utilization are considered. Both zones have an identical number of channels and their spectrum access is modeled according to the simple model discussed in Section 3.3.1. The arrival rates differ and are denoted by $\lambda_{PU,1}$ and $\lambda_{PU,2}$. This results in a different amount of available channels for the SU system in each zone denoted by i and j . It is further assumed that the SU network is located at the fringe of both zones, that is, part of the SUs detects the first PU system, the other part detects the second one. For the communication between the two groups of SUs only those channels can be used that are not occupied in neither PU zone. The number of the common channels is denoted by κ and is derived in the following.

With the arrival rates $\lambda_{PU,1}$ and $\lambda_{PU,2}$ and the service rates $\mu_{PU,1}$ and $\mu_{PU,2}$ the probabilities for the number of available channels in each spectrum occupation zone can be derived with (3.17) and (3.20). Figure 3.16 shows the probabilities for the number of available channels for two PU systems with different arrival rates, with $\rho = \frac{\lambda}{\mu}$. Assuming that i and j are independent the possible number of in common available channels κ depending on i and j is visualized in the matrix shown

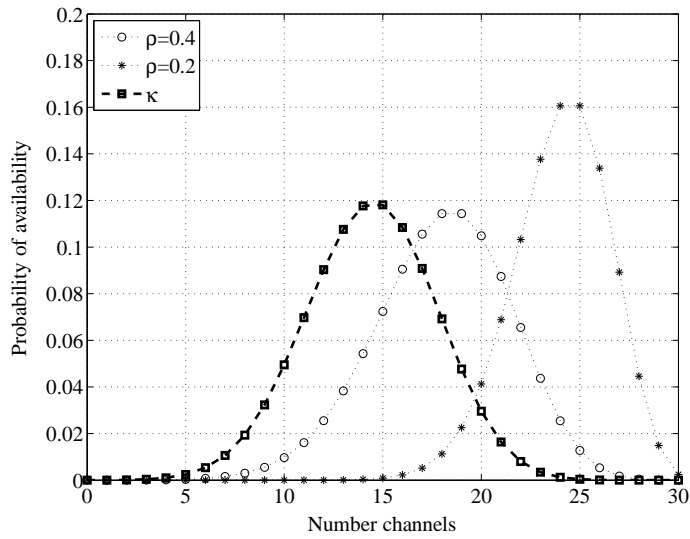


Figure 3.16 Probabilities for the number of available channels for two PU systems with different arrival rates.

in Figure 3.17. Even if both occupation zones have for example two unused channels, they are only available to the SU system if they are identical. This results in the possibility that either both, none or only one channel is available in both zones regarding the given example. Thus, κ follows a hyper geometric distribution [72]

$i \backslash j$	0	1	2	3	4
0	0	0	0	0	0
1	0	0, 1	0, 1	0, 1	1
2	0	0, 1	0, 1, 2	1, 2	2
3	0	0, 1	1, 2	2, 3	3
4	0	1	2	3	4

Figure 3.17 Possible numbers of in common available channels depending on the number of available channels in both occupation zones.

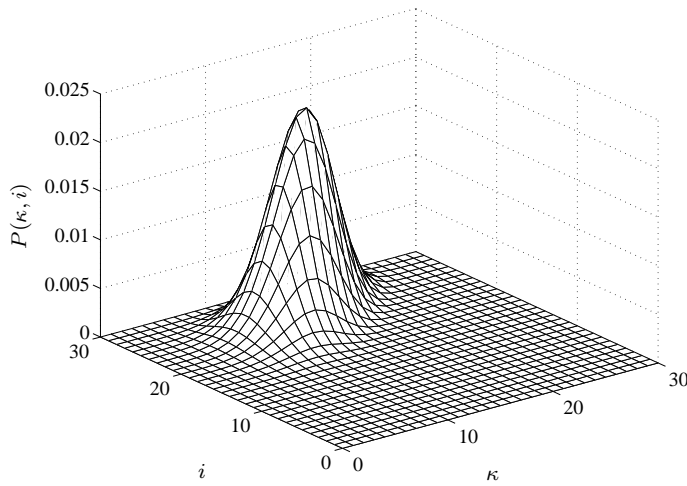


Figure 3.18 Probability of the number of overall available channels κ depending on the number of available channels in one spectrum occupation zone.

with the pdf

$$P(\kappa, i, j) = \frac{\binom{i}{\kappa} \cdot \binom{k-i}{j-\kappa}}{\binom{k}{j}} \quad \text{with} \quad \begin{array}{l} 0 \leq \kappa \leq j \\ \kappa \leq i \\ j - \kappa \leq k - i, \end{array} \quad (3.32)$$

and can be represented by the three-dimensional matrix $\mathbf{A} \in \mathbb{R}^3$ with the elements $a_{i,j,\kappa}$. Figure 3.18 shows the resulting probabilities independent from j , that is, for

$$P(\kappa, i) = \sum_j p_j a_{i,j,\kappa}. \quad (3.33)$$

Finally, $P(\kappa)$ is given by

$$P(\kappa) = \sum_i \sum_j p_i p_j a_{i,j,\kappa} \quad (3.34)$$

and is plotted in Figure 3.16 for the assumed arrival rates and service rates. As expected, the expectation of $P(\kappa)$ is smaller than the expectations of both single expectations of the occupation zones.

4 Overlay Systems in Ad Hoc Mode

In the previous chapter, general aspects and challenges for the coexistence of a SU system in the same frequency band as a PU system were discussed. For operating an SU system in ad hoc mode, several additional challenges have to be met, due to the special properties of ad hoc networks. In this chapter, the properties of ad hoc networks and overlay systems are addressed jointly and the necessary modifications regarding the detection of PU systems are discussed. First, a short introduction on wireless ad hoc networks is given, followed by a graph theory model. Then the concept of distributed detection is applied to ad hoc networks and a distributed approach for the signaling of the detection results is proposed. The discussed distributed detection approach requires that all SUs are silent during the detection phase, so that the detection results are not interfered by transmissions of the SU system. Therefore, finally also an approach for the coordination of the detection phases in the MAC layer is proposed.

4.1 Ad Hoc Networks

Ad hoc networks are networks without infrastructure, that is, there is no central unit for the coordination of the medium access or other managing tasks. All organization and coordination has to be done by the network in a distributed manner, since in the general case, all stations are considered to have equal functionality. The first ad hoc networks were developed by the Defense Advanced Research Projects Agency (DARPA) in the 1970s and were called "packet radios" [52]. Meanwhile, there are many different applications and scenarios for networks without or with only little infrastructure, and especially topics like sensor networks and mobile ad hoc networks (MANETs) are in the focus of research.

Applications range from small scale networks, for example private area networks (PANs) with only few nodes, to large scale networks, like sensor networks, with hundreds or even thousands of nodes. One standard providing ad hoc functionality for PANs operating in the 2.4 GHz ISM band is Bluetooth [5], which dates back to 1997 and is used for setting up communication links for example between mobile phones and wireless head sets, or a computer and a printer. Bluetooth can also be considered as a wireless replacement of the serial cable. Another large area for civil applications of ad hoc networks is the field of emergency operations. In dis-

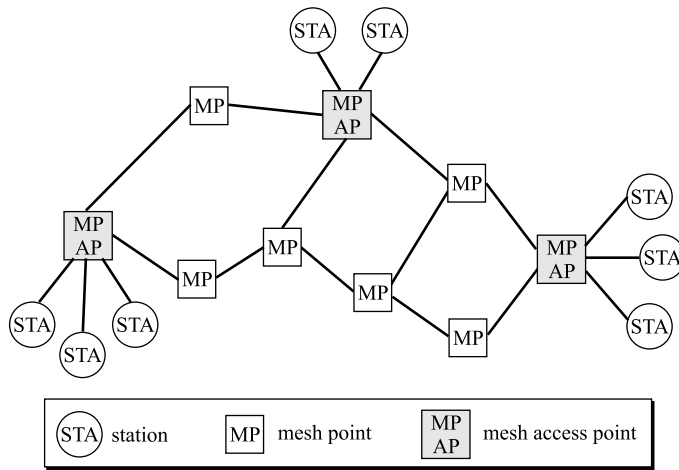


Figure 4.1 Mesh network.

aster scenarios, as for example in hurricane, flood, fire or search/rescue situations, the emergency services rely on being able to communicate as soon as possible, for which setting up an infrastructure based communication system may not be fast enough. Ad hoc networks are also ideally suited for communication between traffic participants, for example car-to-car communication also called vehicular ad hoc networks (VANETs) [47], or aeronautical communication. Technological advances in producing small and smart devices enable the concept of sensor networks [8, 37]. A large number of inexpensive sensor devices is deployed in an area of interest, where they are used for the collection of environmental data, for example measuring toxicity levels at hazardous sites or weather forecasting. After their deployment, the sensor devices form a mesh network and collaborate for gathering the data at an access point which serves as an interface for accessing the sensor data. Mesh networks [9, 30] are multi-hop ad hoc networks where some of the nodes are connected to the infrastructure, but all nodes in the network can use the available services through dynamic multi-hop connections. Finally, ad hoc networks are definitely of interest for military applications.

The main advantages of ad hoc networks are their fast and easy deployment and the decreased dependence on infrastructure. They therefore enable networking anywhere and anytime, and thus are typically deployed in places where it is difficult to set up infrastructural support. The resulting challenges include an increased demand of resources needed for signaling network control information, scalability,

address assignment, routing, security issues, and also an intelligent power budget management for maximizing the life-time of ad hoc nodes, since they are typically powered by a battery.

Communication standards supporting a mesh or ad hoc mode are for example WiMAX (IEEE 802.16 [4]), IEEE 802.11 (mesh mode: 802.11s) [6], terrestrial trunked radio (TETRA) [3] or Bluetooth.

4.2 Model Based on Graph Theory

Graph theory is a suitable tool to describe and analyze the properties of ad hoc networks. An ad hoc network consists of N_{SU} SUs which are distributed in space and are represented by the set of nodes $\mathcal{V} = \{v_1, v_2, \dots, v_{N_{\text{SU}}}\}$ with a cardinality of $|\mathcal{V}| = N_{\text{SU}}$. To indicate the positions of the nodes the vectorial notation $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_{\text{SU}}}\} = \mathcal{V}_{\text{pos}}$ is used. In this work only the two-dimensional space \mathbb{R}^2 is considered, so that $\mathcal{V}_{\text{pos}} \subset \mathbb{R}^2$. If two SUs are able to establish a communication link, they are connected by an edge in the corresponding graph. In general graph theory, each node may be connected with any other node of the network [24]. In contrast, for a wireless ad hoc network the edges are determined by the positions of the nodes and especially depend on the transmission ranges of the SUs. To ensure a reliable reception of the signals at the receiver, a minimum SNR is required which depends on the applied transmission mode. Therefore also the signal power at the receiver has to be above a certain threshold. The area in which the signal power of an SU is high enough for a reliable reception is called *coverage area* of the SU. For modeling the coverage area in the context of ad hoc networks, often a model for the signal power is used which abstracts many properties of the environment and concludes them in a simple way [67]:

$$P_r(r_0) \propto \frac{P_t}{r_0^\alpha}. \quad (4.1)$$

The receiver power P_r is proportional to the transmitter power P_t and reciprocally proportional to the distance r_0 between transmitter and receiver. The properties of the environment are summarized in the transmission coefficient α . Typical values for α are $2 \leq \alpha \leq 6$. The simplified coverage area of a node results in a disc centered around the node with a radius proportional to $\sqrt[\alpha]{P_t}$.

Assuming that all SUs are identically equipped, that is, they have the same transmission power and thus also the same transmission range, the resulting network topology can be completely described by the positions of the SUs and the system-

wide transmission range r_0 . Accordingly, an ad hoc network can be represented by an undirected graph $G = G(\mathcal{V}, r_0)$, in which two nodes $v_1, v_2 \in \mathcal{V}$ are connected by an edge if they are within each others transmission range, that is, $\|\mathbf{v}_1 - \mathbf{v}_2\| \leq r_0$ where $\|\cdot\|$ denotes the Euclidian norm on \mathbb{R}^2 . A graph is called *connected* if there exists a path between every pair of nodes. Whether a graph is connected or not can for example be determined with the help of spectral graph theory [35]. The critical transmission range (CTR) of a graph is defined as the minimum transmission range still guaranteeing a fully connected graph. The *neighborhood* of a node $v \in \mathcal{V}$ is defined as the set of nodes adjacent to v and is denoted by $\Gamma_G(v)$. The *degree* of v is defined as its number of neighbors $d(v) = |\Gamma_G(v)|$ and if $d(v) = 0$, node v is called *isolated*.

In general, it is more interesting to achieve conclusions not only about a single ad hoc network, but to analyze and evaluate the behavior of complete classes of networks. For this reason, random graphs are investigated. Graphs, in which the positions of the nodes are randomly distributed and the edges depend on the SUs' transmission ranges are also called random geometric graphs [65].

4.2.1 Number of Neighbors

The number of neighbors $d(v)$ of an arbitrary node $v \in \mathcal{V}$ in an ad hoc network plays an important role for the further investigations in section 4.3.2. Therefore, the relevant basics of graph theory in this context are discussed first. The number $d(v)$ depends on the distribution of the nodes in the system area \mathbf{A} of size A and thus is a random variable D . Hence, a Cartesian coordinate system is considered in which the two-dimensional random variable $\mathbf{V} = [V_1, V_2]$ describes the positions of the nodes on the abscissa and ordinate. In general, \mathbf{V} can be distributed arbitrarily. Depending on the size and topology of the system area, it can be interesting to either consider the effects at the border of \mathbf{A} , or to neglect them. When looking for example at a very large area with a high node density, the area near the border only plays an insignificant role for the behavior in the inside of the system area. Since border effects will always degrade but never increase the overall performance regarding distributed detection, the borderless scenario will serve as a benchmark.

No Border Effects

A scenario without border effects can be derived by considering a network with an infinite system area, that is, $A \rightarrow \infty$ and $N_{\text{SU}} \rightarrow \infty$ with a constant node density $\rho = \frac{N_{\text{SU}}}{A}$. Under these terms, the number of nodes in every finite subarea follows a

Poisson distribution and the number of nodes in disjoint subareas are independent random variables [38], that is, we have a homogeneous Poisson point process with intensity ρ for the number of nodes. Since $d(v)$ is equivalent to the number of nodes placed in a disk of radius r_0 centered around v , the number of neighbors also follows a Poisson distribution and has to be described by the random variable D . The probability that v has d neighbors is then given by

$$P_\infty(D = d) = P_\infty(d | \mu) = \frac{\mu^d}{d!} e^{-\mu} \quad (4.2)$$

with expectation

$$E\{D\} = \mu = \rho\pi r_0^2, \quad (4.3)$$

the index ∞ referring to the scenario neglecting border effects. Due to the homogeneity of the Poisson process, μ is independent of the location and therefore identical for every node.

With Border Effects

Let \mathbf{A} now be a disk with radius a centered around the origin in which N_{SU} SUs are distributed. The nodes' normalized transmission range is defined as $\hat{r}_0 = r_0/a$ and $\hat{r} = r/a$ is the normalized radial component of the node position when using polar coordinates with $r = \|\mathbf{v}\|$. Figure 4.2 shows an example scenario with $N_{\text{SU}} = 250$ SUs and $\hat{r}_0 = 0.2$. Let $P_0(v)$ be the probability that a second node is placed within the transmission range of the given node v . Note that $P_0(v)$ depends on the location of v due to the bounded system area, and therefore $P_0(v) = P_0(\mathbf{v})$. The probability that $D = d$ nodes are placed within the transmission range of v (that is, v has degree d), can then be described by a binomial distribution:

$$P_\odot(D = d | \mathbf{v}) = \binom{N_{\text{SU}} - 1}{d} P_0(\mathbf{v})^d (1 - P_0(\mathbf{v}))^{N_{\text{SU}} - d - 1}. \quad (4.4)$$

For large N_{SU} and small $P_0(v)$, the probability $P_\odot(D = d | \mathbf{v})$ can be approximated by a Poisson distribution with the location dependent expectation $\mu(v) = (N_{\text{SU}} - 1)P_0(\mathbf{v})$. The unconditional probability for the number of neighbors in a network can be derived by integration over all possible node positions:

$$P_\odot(D = d) = P_\odot(d | N_{\text{SU}}, r_0) = \iint_{\mathbf{A}} P_\odot(D = d | \mathbf{v}) f_{\mathbf{V}}(\mathbf{v}) d\mathbf{v} \quad (4.5)$$

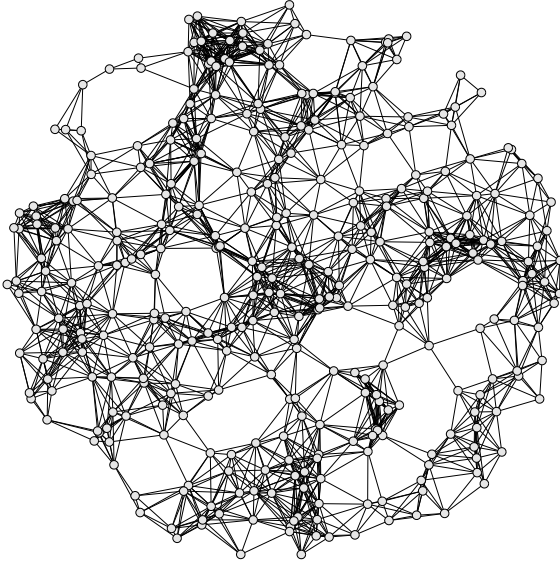


Figure 4.2 Ad hoc network with $N_{\text{SU}} = 250$ SUs with a normalized transmission range $\hat{r} = 0.2$.

which represents the result for an arbitrary node distribution with a pdf $f_{\mathbf{V}}(\mathbf{v})$ for \mathbf{V} in the area \mathbf{A} . The index \odot refers to the scenario with border effects. For the following investigations a uniform distribution with the pdf

$$f_{\mathbf{V}}(\mathbf{v}) = \begin{cases} \frac{1}{A} & \text{for } \mathbf{v} \in \mathbf{A} \\ 0 & \text{otherwise.} \end{cases} \quad (4.6)$$

is assumed. Using polar coordinates (with the center of \mathbf{A} as origin), $P_0(\mathbf{v})$ especially depends on the radial component, that is, $P_0(\hat{r})$, since it determines the distance to the border of the system area. Based on geometrical considerations, $P_0(\hat{r})$ is given by [23]:

$$P_0(\hat{r}) = \begin{cases} \hat{r}_0^2 & \text{for } 0 \leq \hat{r} \leq 1 - \hat{r}_0 \\ \frac{1}{\pi} \left(\hat{r}_0^2 \arccos \frac{\hat{r}^2 + \hat{r}_0^2 - 1}{2\hat{r}\hat{r}_0} \right. \\ \quad \left. + \arccos \frac{\hat{r}^2 - \hat{r}_0^2 + 1}{2\hat{r}} - \frac{1}{2}\sqrt{\zeta} \right) & \text{for } 1 - \hat{r}_0 < \hat{r} \leq 1 \end{cases} \quad (4.7)$$

with $\zeta = (\hat{r} + \hat{r}_0 + 1)(-\hat{r} + \hat{r}_0 + 1)(\hat{r} - \hat{r}_0 + 1)(\hat{r} + \hat{r}_0 - 1)$. The first case of (4.7) gives the probability when the transmission range of a node is completely in the system area. The probability for the fringe, that is, the region where only part of the transmission range of v is in the system area ($1 - \hat{r}_0 < \hat{r} \leq 1$), is given in the second case.

By combining, the location independent probability for the number of neighbors assuming a uniform node distribution can be derived. For this, (4.4) and (4.6) are inserted in (4.5), yielding:

$$P_{\odot}(d | N_{\text{SU}}, r_0) = \int_0^1 2\hat{r} \cdot \binom{N_{\text{SU}} - 1}{d} P_0(\hat{r})^d (1 - P_0(\hat{r}))^{N_{\text{SU}} - d - 1} d\hat{r}. \quad (4.8)$$

4.3 Distributed Detection in Ad Hoc Networks

The detection subsystem of an overlay system is a vital component for the coexistence of a PU and SU system in the same frequency band. A reliable detection of the PU system is necessary for avoiding collisions with the PU system and reducing mutual interference as well as for maximizing the utilization of spectrum holes. The detection probability P_{dct} and false alarm probability P_{false} cannot be optimized independently, leading to a tradeoff between a high reliability regarding the detection and an optimum utilization of the idle spectrum parts. Since the detection of the PU system has high priority, in an overlay scenario the required detection probability is given (for example 99.9%) and the false alarm probability is then determined by the applied detector model. Results of previous work [83] show that the performance of the detection subsystem in a single SU is not sufficient to efficiently operate an overlay system. According to the derived receiver operating characteristics (ROCs), the achieved false alarm probabilities are too high for the required detection probability of 99.9%. One approach to face this challenge is the concept of distributed detection [83], where the detection results of several SUs are combined and therefore a better system-wide ROC can be achieved.

In this section, first the main aspects of distributed detection and its application in an overlay system operating in infrastructure mode with an access point, as well as the basics of the used detector model are summarized. Then a new concept is proposed, which transfers the distributed detection approach to an ad hoc scenario. The properties of an ad hoc network related to the performance of the detection subsystem are investigated and it is discussed how a required detection probability can be achieved in the context of an ad hoc scenario. Furthermore, especially

the signaling mechanisms have to be modified: An approach is proposed that efficiently combines distributed detection and signaling among the neighbors of the SUs operating in ad hoc mode.

4.3.1 Distributed Detection

The overall detection performance of an overlay system can be significantly increased by applying a distributed detection approach where several SUs of the system combine their single detection results. Based on the detection probability P_{dct} and false alarm probability P_{false} of a single SU, and assuming that L SUs are involved in the distributed detection, the resulting detection probability $P_{\text{dct}}^{\text{dist}}$ and false alarm probability $P_{\text{false}}^{\text{dist}}$ are given by [83]

$$\begin{aligned} P_{\text{dct}}^{\text{dist}}(L) &= 1 - (1 - P_{\text{dct}})^L \\ P_{\text{false}}^{\text{dist}}(L) &= 1 - (1 - P_{\text{false}})^L. \end{aligned} \quad (4.9)$$

Although the false alarm probability is also increased by the distributed approach, it is shown that this effect is overcompensated due to the convexity of the ROC and that the overall detection performance thus is increased significantly, which allows an efficient operation of an overlay system.

Equation (4.9) implies that the single detection results have to be combined and processed at some place and hence also some signaling has to take place. In case of an overlay system operating in infrastructure mode, an efficient approach for signaling the single detection results of the contributing SUs to the SU-AP, called *boosting protocol*, is proposed in [82]. The detection period is initiated by the SU-AP, for example by sending a special control packet, and is organized in the three phases *detection*, *collection* and *broadcast*, as shown in Figure 4.3. The main idea of this approach can be summarized as follows:

- In the detection phase (phase A in Figure 4.3) all SUs are silent and perform a detection.
- When detecting a new PU, an SU transmits an unmodulated boosting signal in the collection phase (phase B) on those subcarriers which are newly occupied by the PU. The SU-AP does not transmit a boosting signal and instead performs another detection phase. The boosting signals will of course increase the interference towards the PU system for a short moment, but this additional signal will also increase the detection probability at the access

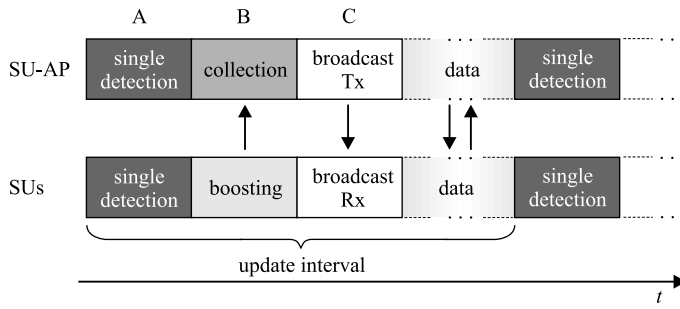


Figure 4.3 Detection period containing three phases for distributed detection in infrastructure mode.

point, where the boosting signals of all SUs superimpose and thus inherently perform an OR-operation without any additional processing overhead. Note that the boosting signal is sent only once, when the PU access is newly detected. The resulting interference towards the PU system is therefore limited to one detection period. The protocol proposed in [82] also handles the signaling for the situation where a channel previously allocated by the PU system becomes available for the SU system. This signaling scheme can also be adopted for the ad hoc scenario.

- At the access point a system-wide spectrum occupation vector is derived based on the combination of the single user detection results, which is then distributed (phase C) within the complete cell and is mandatory for all associated SUs until an update is broadcasted.

This cycle based on the three phases is repeated every update interval.

Detection Probability and False Alarm Probability

Since in the detection phase only a single user detection is performed where no boosting signal is present, the resulting detection and false alarm probability for the detection phase can be determined with (3.14) and (3.15), respectively.

In the collection phase the SUs transmit a boosting signal and only the SU-AP performs a further detection phase. Due to the additional boosting signal at the receiver, the detector model described in Section 3.2.1 must be extended by the boosting signal component as shown in [81] to derive the detection and false alarm probability for the collection phase at the SU-AP.

Introducing the boosting signal $b(k)$ —which is assumed to be zero-mean Gaussian distributed—the received signal is now

$$r(k) = s(k) + b(k) + n(k). \quad (4.10)$$

Accordingly, also (3.10) has to be extended, yielding:

$$z_m = R(m) = S(m) + B(m) + N(m) \quad m = 1, 2, \dots, M \quad (4.11)$$

where $\{B(m) = B_x(m) + jB_y(m)\}_{m=1}^M$ is the FFT of $\{b(k)\}_{k=1}^M$. The mean power of $B(m)$ is denoted by $2\sigma_B^2$. Assuming that all $B(m)$, $m = 1, 2, \dots, M$; are independent, similar to (3.13) the conditional pdf $f_{\Lambda|W, \text{PU}}(\lambda | w, \text{PU})$ with W representing the number of simultaneously boosting SUs is given by:

$$f_{\Lambda|W, \text{PU}}(\lambda | w, \text{PU}) = \begin{cases} \frac{\lambda^{M-1}}{2^M (\sigma_S^2 + w\sigma_B^2 + \sigma_N^2)^M (M-1)!} & \text{for } \lambda \geq 0 \\ \cdot \exp\left(-\frac{\lambda}{2(\sigma_S^2 + w\sigma_B^2 + \sigma_N^2)}\right) & \\ 0 & \text{for } \lambda < 0. \end{cases} \quad (4.12)$$

Since an SU only sends a boosting signal when it previously detected a PU in the detection phase, not always all SUs in the system contribute to the access point's detection. The number W of simultaneously received boosting signals at the access point follows a binomial distribution with the single detection probability P_{dct} from (3.14):

$$P(W = w) = \binom{L}{w} (P_{\text{dct}})^w (1 - P_{\text{dct}})^{L-w}, \quad 0 \leq w \leq W. \quad (4.13)$$

The conditional pdfs $f_{\Lambda|\text{PU}}(\lambda | \text{PU})$ and $f_{\Lambda|\overline{\text{PU}}}(\lambda | \overline{\text{PU}})$ for the detection at the SU-AP are then given by [81]:

$$f_{\Lambda|\text{PU}}(\lambda | \text{PU}) = \sum_{w=0}^L \binom{L}{w} (P_{\text{dct}})^w (1 - P_{\text{dct}})^{L-w} \cdot f_{\Lambda|W, \text{PU}}(\lambda | w, \text{PU}). \quad (4.14)$$

and

$$f_{\Lambda|\overline{\text{PU}}}(\lambda | \overline{\text{PU}}) = \sum_{w=0}^L \binom{L}{w} P_{\text{false}}^w (1 - P_{\text{false}})^{L-w} \cdot f_{\Lambda|W, \overline{\text{PU}}}(\lambda | w, \overline{\text{PU}}). \quad (4.15)$$

The pdf $f_{\Lambda|W,\overline{\text{PU}}}(\lambda | w, \overline{\text{PU}})$ can be obtained from $f_{\Lambda|W,\text{PU}}(\lambda | w, \text{PU})$ by setting $\sigma_S^2 = 0$. Similar to the single user detection probability the calculation of the system's detection probability $P_{\text{dct}}^{\text{cell}}(L)$ and false alarm probability $P_{\text{false}}^{\text{cell}}(L)$ in the cell can now be performed by integration:

$$\begin{aligned} P_{\text{dct}}^{\text{cell}}(L) &= \int_{\lambda_2}^{\infty} \sum_{w=0}^L \binom{L}{w} (P_{\text{dct}})^w (1 - P_{\text{dct}})^{L-w} \cdot f_{\Lambda|W,\text{PU}}(\lambda | w, \text{PU}) d\lambda \\ &= \sum_{w=0}^L \binom{L}{w} (P_{\text{dct}})^w (1 - P_{\text{dct}})^{L-w} \cdot P_{\text{dct}}(w) \end{aligned} \quad (4.16)$$

and

$$\begin{aligned} P_{\text{false}}^{\text{cell}}(L) &= \int_{\lambda_2}^{\infty} \sum_{w=0}^L \binom{L}{w} (P_{\text{false}})^w (1 - P_{\text{false}})^{L-w} \cdot f_{\Lambda|W,\overline{\text{PU}}}(\lambda | w, \overline{\text{PU}}) d\lambda \\ &= \sum_{w=0}^L \binom{L}{w} (P_{\text{false}})^w (1 - P_{\text{false}})^{L-w} \cdot P_{\text{false}}(w) \end{aligned} \quad (4.17)$$

where λ_2 describes the threshold in the access point during the collection phase. Note that λ_2 differs from λ_1 . To enable a clear presentation in the following sections, the detection probability and false alarm probability depending on the number of boosting SUs introduced as

$$P_{\text{dct}}(w) = \int_{\lambda_2}^{\infty} f_{\Lambda|W,\text{PU}}(\lambda | w, \text{PU}) d\lambda \quad (4.18)$$

and

$$P_{\text{false}}(w) = \int_{\lambda_2}^{\infty} f_{\Lambda|W,\overline{\text{PU}}}(\lambda | w, \overline{\text{PU}}) d\lambda. \quad (4.19)$$

Note that $P_{\text{dct}}(w) = P_{\text{dct}}$ for $w = 0$. Inserting (4.12) in (4.16) and solving the

integral yields

$$\begin{aligned}
P_{\text{dct}}^{\text{cell}}(L) &= \sum_{w=0}^L \binom{L}{w} (P_{\text{dct}})^w (1 - P_{\text{dct}})^{L-w} \\
&\cdot \frac{1}{2^M (\sigma_S^2 + w\sigma_B^2 + \sigma_N^2)^M (M-1)!} \\
&\cdot \exp \left\{ -\frac{\lambda_2}{2(\sigma_S^2 + w\sigma_B^2 + \sigma_N^2)} \right\} \\
&\cdot \left(2(\sigma_S^2 + w\sigma_B^2 + \sigma_N^2) \lambda_2^{M-1} \right. \\
&\left. + \sum_{m=1}^{M-1} (2(\sigma_S^2 + w\sigma_B^2 + \sigma_N^2))^{m+1} \lambda_2^{M-m-1} \prod_{l=1}^m (M-l) \right).
\end{aligned} \tag{4.20}$$

The false alarm probability in a cell can be calculated in a similar way and yields

$$\begin{aligned}
P_{\text{false}}^{\text{cell}}(L) &= \sum_{w=0}^L \binom{L}{w} (P_{\text{false}})^w (1 - P_{\text{false}})^{L-w} \\
&\cdot \frac{1}{2^M (w\sigma_B^2 + \sigma_N^2)^M (M-1)!} \\
&\cdot \exp \left\{ -\frac{\lambda_2}{2(w\sigma_B^2 + \sigma_N^2)} \right\} \cdot \left(2(w\sigma_B^2 + \sigma_N^2) \lambda_2^{M-1} \right. \\
&\left. + \sum_{m=1}^{M-1} (2(w\sigma_B^2 + \sigma_N^2))^{m+1} \lambda_2^{M-m-1} \prod_{l=1}^m (M-l) \right).
\end{aligned} \tag{4.21}$$

4.3.2 Theoretical Bounds

When applying the distributed detection approach to an ad hoc scenario, the interesting question arises how the overall detection performance of the network is related to the properties of the ad hoc network. When looking for example at a scenario where a sensor network operating as an overlay system shall be deployed in a well defined area, it is important to know how many nodes have to be distributed in order to achieve a detection probability still meeting the requirements of the PU system. Similar to the cell-based distributed detection probability $P_{\text{dct}}^{\text{cell}}(L)$ and false alarm probability $P_{\text{false}}^{\text{cell}}(L)$ in an infrastructure scenario, now the network-based detection probabilities $P_{\text{dct}}^{\text{net}\infty}(\rho, r_0)$ and $P_{\text{dct}}^{\text{net}}(N_{\text{SU}}, r_0)$ —for the borderless scenario and the scenario considering border effects, respectively—are defined for

an overlay system operating in ad hoc mode [21]. Therefore, the results from graph theory (Section 4.2) are combined with the detection and false alarm probability for distributed detection (Section 4.3.1) to give a theoretical bound for the minimum necessary number of randomly distributed nodes in an ad hoc network, still achieving the required average detection probability. Since this is a lower bound, it is reasonable to use the best case detector model as mentioned in Section 3.2.1.

In contrast to the infrastructure based scenario, where the SU-AP knows how many SUs are contributing to the distributed detection process and generates the cell-based spectrum occupation vector, in an ad hoc scenario no central processing SU is available. Nevertheless, each SU can enhance its detection performance by applying the distributed detection approach based on the single detection results of its neighbors. Signaling issues are neglected at this point and are discussed in Section 4.3.3. Assuming that an SU knows the number of its neighbors and that every neighbor contributes to the distributed detection process, the network detection probability and network false alarm probability are defined as

$$\begin{aligned} P_{\text{dct}}^{\text{net}} &= \sum_{d=0}^{N_{\text{SU}}-1} P(d) \cdot P_{\text{dct}}^{\text{dist}}(L) \\ P_{\text{false}}^{\text{net}} &= \sum_{d=0}^{N_{\text{SU}}-1} P(d) \cdot P_{\text{false}}^{\text{dist}}(L). \end{aligned} \quad (4.22)$$

$P_{\text{dct}}^{\text{dist}}(L)$ and $P_{\text{false}}^{\text{dist}}(L)$ are weighted by the probability that a node v has d neighbors. $P(d)$ denotes the general probability that a SU has d neighbors and depends on the considered scenario. Since the number of neighbors d excludes the node itself, which nevertheless also performs a single user detection, the total number of contributing nodes is always $L = d + 1$. For clarity reasons, the following derivations are only performed for the detection probabilities. The results for the false alarm probability can be easily derived by substituting $P_{\text{dct}}^{\text{net}}$ with $P_{\text{false}}^{\text{net}}$, unless otherwise noted.

No Border Effects

In the borderless scenario, according to (4.2) the probability that a node has d neighbors is given by $P(d) = P_{\infty}(d, \mu)$, so that $P_{\text{dct}}^{\text{net}\infty}(\rho, r_0)$ depends on the node density and the transmission range, and thus with (4.3) also on μ :

$$P_{\text{dct}}^{\text{net}\infty}(\mu) = \sum_{d=0}^{\infty} P_{\infty}(d, \mu) \cdot P_{\text{dct}}^{\text{dist}}(d + 1). \quad (4.23)$$

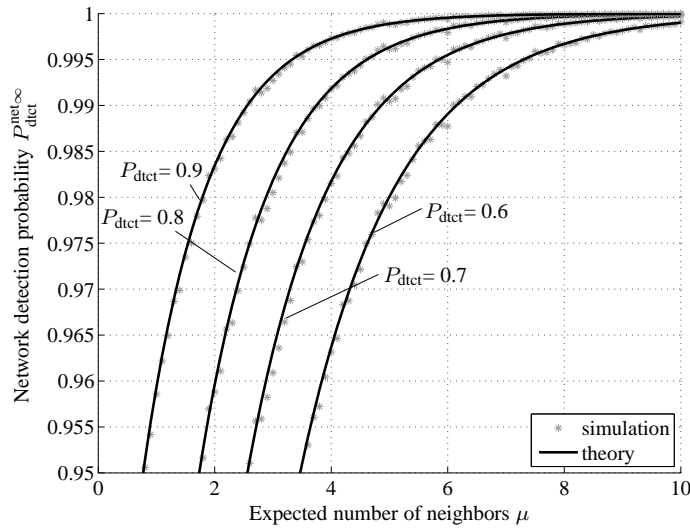


Figure 4.4 Network detection probability $P_{dtct}^{net,∞}$ in the borderless scenario for various single detection probabilities P_{dtct} depending on the expected number of neighbors μ .

Plugging (4.2) and the first equation of (4.9) into (4.23) and with $L = d + 1$ the detection probability is given by

$$P_{dtct}^{net,∞}(\mu) = \sum_{d=0}^{\infty} \frac{\mu^d}{d!} e^{-\mu} \cdot (1 - (1 - P_{dtct})^{d+1}). \quad (4.24)$$

In a similar way, the SU system's network false alarm probability $P_{false}^{net,∞}(\mu)$ can be expressed when using the second equation of (4.9):

$$P_{false}^{net,∞}(\mu) = \sum_{d=0}^{\infty} \frac{\mu^d}{d!} e^{-\mu} \cdot (1 - (1 - P_{false})^{d+1}). \quad (4.25)$$

Figure 4.4 shows the plots of $P_{dtct}^{net,∞}(\mu)$ for several single detection probabilities P_{dtct} . By forcing a network detection probability $P_{dtct}^{net,∞} = 0.999$ and with for example $P_{dtct} = 0.8$ the expected number of neighbors needs to be at least $\mu = 6.6$. According to (4.3), the required μ can be achieved by adjusting either the transmission range of the nodes or the node density. To simulate the borderless scenario, the SUs were placed randomly in an area sufficiently larger than the system area,

μ	P_{dtct}	P_{false}	$P_{\text{false}}^{\text{net}\infty}$
0	0.999	0.853	0.853
3	0.981	0.384	0.805
5	0.906	0.096	0.441
7	0.781	0.020	0.150
9	0.656	0.005	0.050
19	0.344	≈ 0	0.002

Table 4.1 Ad hoc network's false alarm probability $P_{\text{false}}^{\text{net}\infty}(\mu)$ with $M = 8$, SNR = 2 dB and $P_{\text{dtct}}^{\text{net}\infty} = 0.999$.

but only the SUs in the system area were considered for the results. The number of each SU's neighbors was determined and then the resulting minimum possible distributed detection probability was calculated, assuming that every neighbor contributes to the distributed detection. The simulation results in Figure 4.4 are close to the analytical results.

The effect on the false alarm probability is shown in Table 4.1 for $M = 8$, SNR = 2 dB and a given network detection probability $P_{\text{dtct}}^{\text{net}\infty} = 0.999$. Depending on μ , the required single detection probability for achieving the given network detection probability was calculated with the help of (4.24). The corresponding single false alarm probability is shown in the third column, from which finally the network false alarm probability $P_{\text{false}}^{\text{net}\infty}$ is calculated (Equation (4.25)). The results in Table 4.1 show that an acceptable performance of the detection subsystem can be achieved, when the node density and transmission range of the ad hoc network result in $\mu \geq 7$. Figure 4.5 finally shows the resulting ROCs based on $P_{\text{dtct}}^{\text{net}\infty}$ and $P_{\text{false}}^{\text{net}\infty}$ for different μ compared to the ROC in the single detection case based on P_{dtct} and P_{false} for $M = 4$ and SNR = 0 dB, where M denotes the number of samples stored in the detector's memory as introduced in Section 3.2.1. Note that in this work the axes of the ROCs are swapped compared to the normally used presentation known for example from radar literature. The reason is that for the detection of PU systems usually the detection probability is given instead of the false alarm probability.

With Border Effects

When considering border effects, there is a defined number of SUs in a bounded system area. In this case, the resulting network detection probability does not depend on μ , but on the specific combination of N_{SU} and r_0 , since $P_0(\mathbf{v})$ depends on r_0 according to (4.7). Similar to (4.23), the network detection probability is now

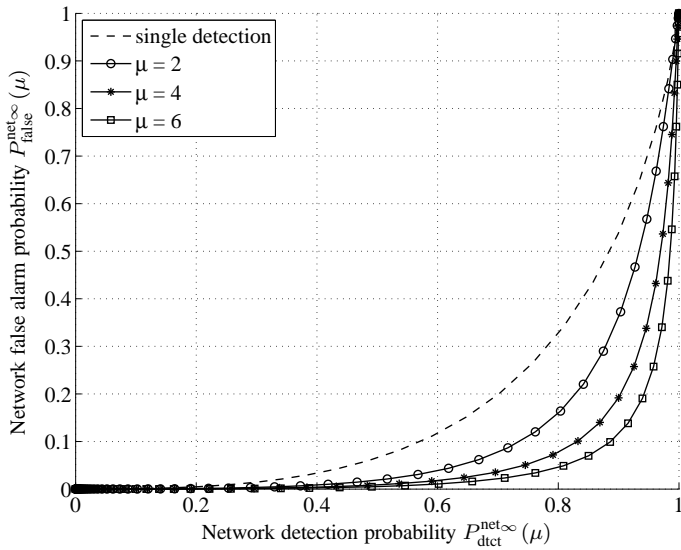


Figure 4.5 ROC for different μ for $M = 4$ and $\text{SNR} = 0$ dB.

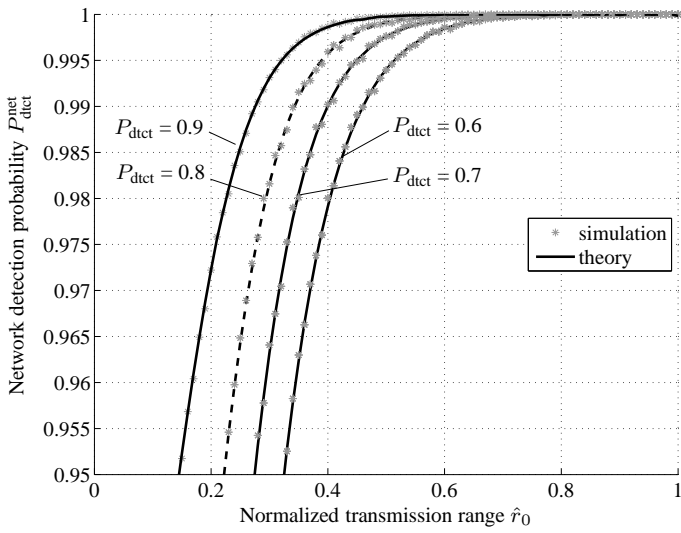


Figure 4.6 Scenario with border effects: network detection probability depending on the normalized transmission range \hat{r}_0 for different single detection probabilities P_{dtct} and $N_{\text{SU}} = 40$.

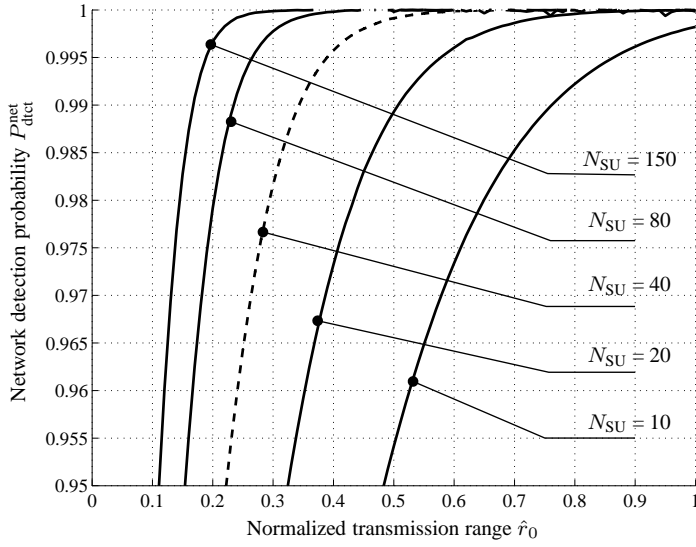


Figure 4.7 Scenario with border effects: network detection probability depending on the normalized transmission range \hat{r}_0 for different numbers of SUs N_{SU} in the system area and $P_{dct} = 0.8$.

given by

$$P_{dct}^{net}(N_{SU}, r_0) = \sum_{d=0}^{N_{SU}-1} P_{\odot}(d | N_{SU}, r_0) \cdot P_{dct}^{dist}(d+1), \quad (4.26)$$

when using (4.8) in (4.22), that is, setting $P(d) = P_{\odot}(d | N_{SU}, r_0)$. This yields

$$P_{dct}^{net}(N_{SU}, r_0) = \sum_{d=0}^{N_{SU}-1} \left(\int_0^1 2\hat{r} \cdot \binom{N_{SU}-1}{d} P_0(\hat{r})^d \cdot (1 - P_0(\hat{r}))^{N_{SU}-d-1} d\hat{r} \cdot (1 - (1 - P_{dct})^{d+1}) \right) \quad (4.27)$$

when plugging (4.8) and the first equation of (4.9) in (4.26). The SU system's network false alarm probability $P_{false}^{net}(N_{SU}, r_0)$ can be derived by replacing P_{dct}^{dist} with P_{false}^{dist} in (4.26).

The plots and simulation results for the detection probabilities depending on the normalized transmission range with $N_{SU} = 40$ are shown in Figure 4.6 for vari-

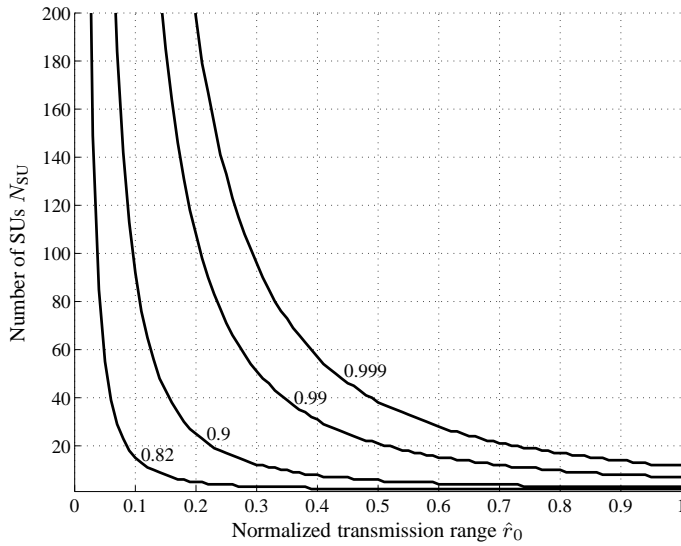


Figure 4.8 With border effects: possible combinations of N_{SU} and \hat{r}_0 resulting in different given network detection probabilities $P_{\text{dct}}^{\text{net}}(N_{\text{SU}}, \hat{r}_0)$ for $P_{\text{dct}} = 0.8$.

ous single detection probabilities. Figure 4.7 shows similar plots, but now with a varying N_{SU} and constant $P_{\text{dct}} = 0.8$. Note that the curves labeled with $N_{\text{SU}} = 40$ and $P_{\text{dct}} = 0.8$ are identical in both Figures. Interpreting Figure 4.7 as a three-dimensional plot with the dimensions N_{SU} , \hat{r}_0 and $P_{\text{dct}}^{\text{net}}$, the contour plots shown in Figure 4.8 can be derived. They display possible combinations of \hat{r}_0 and N_{SU} which yield a given network detection probability. For a given N_{SU} the critical transmission range regarding $P_{\text{dct}}^{\text{net}}$ can be read off. Note that the curves are not smooth, since N_{SU} is an integer number. Finally, Figure 4.9 shows the resulting network ROCs ($\hat{r}_0 = 0.2$, $M = 4$, $\text{SNR} = 0$ dB) for various N_{SU} . The dashed line represents the case for single user detection as a reference. It is obvious that the detection performance increases for an increasing number of SUs in the system area. Note that the values for M and the SNR were chosen in a way that the relative performance gain regarding the single detection case can be seen for increasing N_{SU} . The overall performance can also be increased by using a larger M (resulting in longer detection phases) or by assuming a higher SNR.

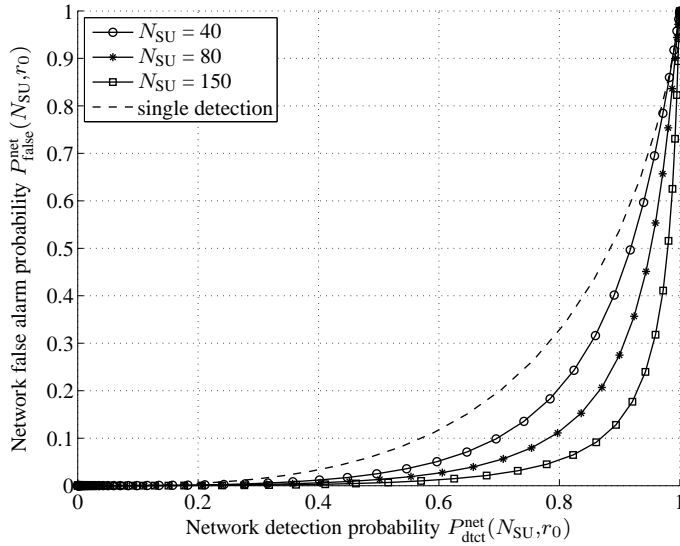


Figure 4.9 Receiver operating characteristic for $\hat{r}_0 = 0.2$, $M = 4$, $\text{SNR} = 0$ dB.

4.3.3 Efficient Signaling of Detection Results

The results regarding the necessary node density derived in Section 4.3.2 are theoretical bounds, assuming that every node can exploit the detection information of all its neighbors. Now a more realistic scenario is considered in which signaling becomes an important issue. The question arises, how ad hoc SUs can exchange their detection results in an efficient way, although there is no central node available that collects the single detection results, derives a system spectrum occupation vector and then distributes it within the complete cell as discussed for the infrastructure based scenario in Section 4.3.1. In an ad hoc scenario all SUs are equal. Therefore, each single SU, or rather the network as the entirety of SUs, must perform all relevant actions of the detection subsystem in a decentralized manner.

In the following, an approach is presented that combines boosting and collection of the single user detection results in a distributed way. Each SU cooperates with its neighbors for detection, thus performing distributed detection in the network. Similar to the infrastructure based case (see Figure 4.3), again a timing structure with three phases is applied for detection and signaling. The detection period is subdivided into the following three slots, as depicted in Figure 4.10:

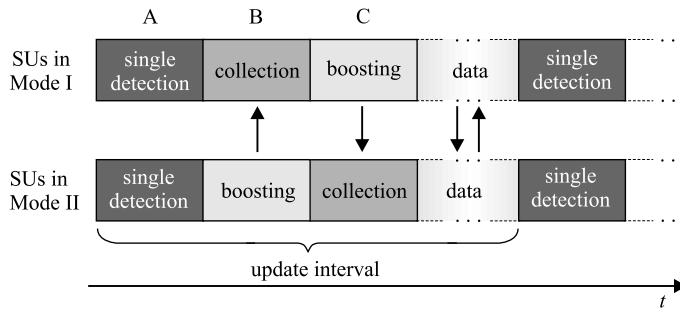


Figure 4.10 Two SUs with complementary boosting phases and a detection period containing three phases for distributed detection in ad hoc mode.

1. The first phase (slot A) is a pure detection phase, where every SU performs a single user detection according to Section 3.2.1 with the predefined detection probability P_{dct} and the corresponding false alarm probability P_{false} .
2. In contrast to the infrastructure-based scenario, every node must perform a boosting *and* a collection phase for exchanging the detection results. Therefore, the original collection phase is duplicated leading to two identical phases (slots B and C) in which two different possible actions (boosting or collection) can be performed. A defined system-wide assignment on which action must be performed in which slot is not reasonable, because this would result in a situation where all SUs are boosting in the same slot and then trying to collect the boosting signals in the other slot. It is more desirable that some of the SUs are boosting while the others are collecting. Therefore, before entering the second phase, each node decides randomly and independently from other SUs whether it will use the second phase (slot B) as collection or boosting phase.
3. The last phase (slot C) will then be used for the complementary action, so an SU either performs the broadcast phase in slot B and the collection phase in slot C or vice versa.

In this way, the probability that all nodes perform the same action in the same slot is kept low. Each SU contributes to the detection of its neighbors as well as it is in turn supported by them. This distributed detection and signaling approach is visualized in Figure 4.11. Finally, the broadcasting phase is omitted for two reasons. First, this would result in a very high signaling effort, since every SU would have to perform the broadcasting phase. Secondly, after performing the collection phase

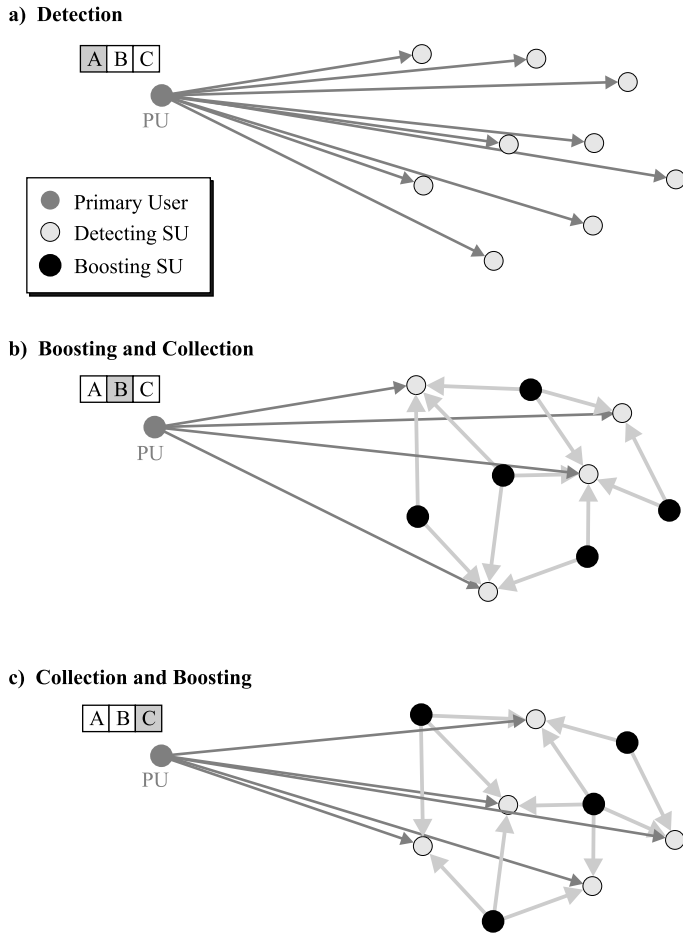


Figure 4.11 Distributed detection and signaling in ad hoc systems.

every SU already has a very good and reliable picture of the PU system’s spectrum occupation, so that the broadcasting phase becomes redundant.

Figure 4.10 also shows the two possible modes for each SU and the interaction of both groups. In this work, the following terminology is used: An SU is referred to be in mode I when first a collection phase is performed followed by a boosting phase. Mode II indicates the complementary sequence. In general, the probability

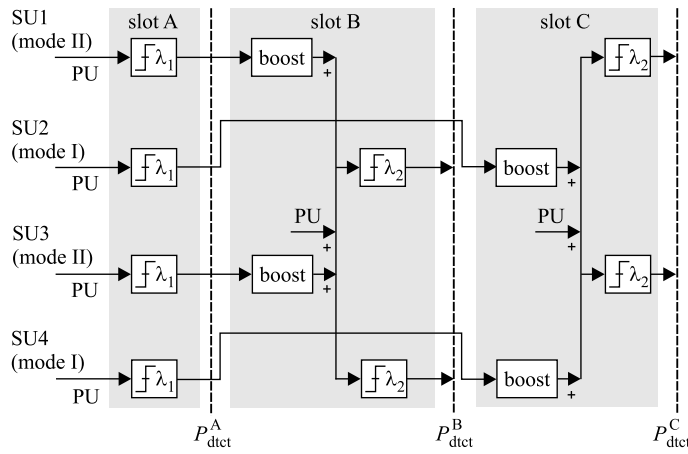


Figure 4.12 Interaction and dependencies of the detectors in different slots and different SUs.

that either slot B or slot C is a boosting phase can be chosen randomly. However, due to the symmetric functionality of both slots as well as for simplicity reasons it is reasonable to set the probability for slot B and slot C either being a boosting phase or collection phase to 0.5. These probabilities are assumed in the further work. Figure 4.12 gives an overview on the structure and the dependencies of the detectors in slot A, B and C. Four SUs are displayed, half of which are operating in mode I (SU2 and SU4) and the other half of which operating in mode II (SU1 and SU3). The detector blocks correspond to Figure 3.6. Assuming that in this example a PU is present, the input of the detectors in slot A only consists of the PU signal, whereas the input of the detectors in slots B and C additionally include the boosting signals of the SUs in the complementary mode. A boosting block only transmits a boosting signal if the connected detector in the preceding slot has detected a PU signal.

In the following, first the resulting detection and false alarm probabilities for slot B and slot C are analyzed and then the combined average network detection and false alarm probabilities are derived for an SU that has completed all three phases.

Detection and False Alarm Probabilities for Slot B

To derive the detection probability $P_{\text{dct}}^{\text{B}}$ in slot B, it is necessary to know the boosting signal sent by the other SUs. A boosting signal is only transmitted, if a PU

signal was detected in slot A and slot B is chosen to be a boosting slot. Therefore, the probability that an SU actually transmits a boosting signal in slot B is given by $0.5 \cdot P_{\text{dctct}}^A$ in case a PU is present and $0.5 \cdot P_{\text{false}}^A$ if no PU is present. P_{dctct}^A and P_{false}^A denote the single user detection and false alarm probabilities in slot A, respectively. Assuming that the SU knows its current number of neighbors d (for example from higher layers), the conditional pdfs can be derived from (4.16) and (4.17) by weighting P_{dctct} and P_{false} with the probability that slot B is used for boosting. Note that in contrast to the theoretical bounds derived in Section 4.3.2 now $L = d$, because now only the neighboring SUs are contributing and not the evaluating SU. The resulting detection probability and false alarm probability in slot B for a known number of neighbors d is then determined by:

$$P_{\text{dctct}}^B(d) = \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{dctct}}^B)^w (1 - 0.5 \cdot P_{\text{dctct}}^B)^{d-w} \cdot P_{\text{dctct}}(w) \quad (4.28)$$

$$P_{\text{false}}^B(d) = \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{false}}^B)^w (1 - 0.5 \cdot P_{\text{false}}^B)^{d-w} \cdot P_{\text{false}}(w) \quad (4.29)$$

with $P_{\text{dctct}}(w)$ and $P_{\text{false}}(w)$ given by (4.18) and (4.19).

Applying the results from Section 4.2 for the borderless scenario as well as the scenario with border effects, the overall average network detection and false alarm probabilities in slot B result in

$$P_{\text{dctct}}^B(\mu) = \sum_{d=0}^{\infty} \frac{\mu^d}{d!} e^{-\mu} \cdot \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{dctct}}^B)^w (1 - 0.5 \cdot P_{\text{dctct}}^B)^{d-w} \cdot P_{\text{dctct}}(w) \quad (4.30)$$

$$P_{\text{false}}^B(\mu) = \sum_{d=0}^{\infty} \frac{\mu^d}{d!} e^{-\mu} \cdot \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{false}}^B)^w (1 - 0.5 \cdot P_{\text{false}}^B)^{d-w} \cdot P_{\text{false}}(w) \quad (4.31)$$

and

$$P_{\text{dctct}}^B(N_{\text{SU}}, r_0) = \sum_{d=0}^{N_{\text{SU}}-1} \int_0^1 2\hat{r} \cdot \binom{N_{\text{SU}}-1}{d} P_0(\hat{r})^d (1 - P_0(\hat{r}))^{N_{\text{SU}}-d-1} \mathbf{d}\hat{r} \quad (4.32)$$

$$\cdot \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{dctct}}^B)^w (1 - 0.5 \cdot P_{\text{dctct}}^B)^{d-w} \cdot P_{\text{dctct}}(w)$$

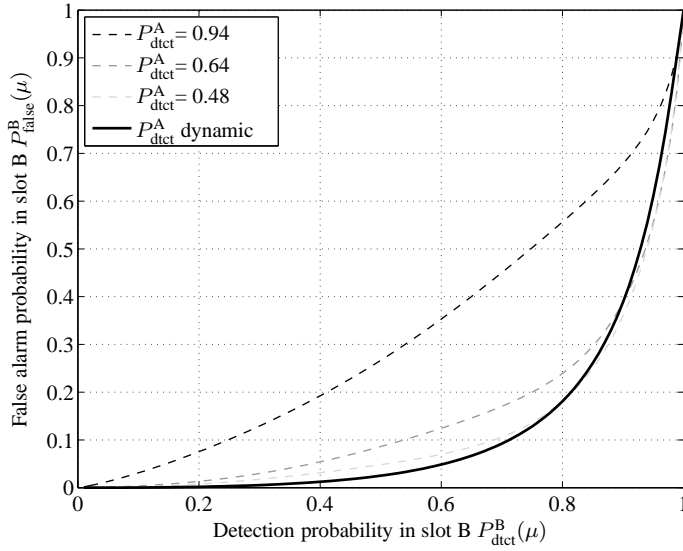


Figure 4.13 ROCs for different fixed single user probabilities compared to the ROC based on dynamically adapted single user probabilities ($M = 4$, $\text{SNR} = 0$ dB, $\text{SNR}_B = 3$ dB, $\mu = 4$).

$$\begin{aligned}
 P_{\text{false}}^B(N_{\text{SU}}, r_0) &= \sum_{d=0}^{N_{\text{SU}}-1} \int_0^1 2\hat{r} \cdot \binom{N_{\text{SU}}-1}{d} P_0(\hat{r})^d (1 - P_0(\hat{r}))^{N_{\text{SU}}-d-1} d\hat{r} \\
 &\quad \cdot \sum_{w=0}^d \binom{d}{w} (0.5 \cdot P_{\text{false}}^B)^w (1 - 0.5 \cdot P_{\text{false}}^B)^{d-w} \cdot P_{\text{false}}(w)
 \end{aligned} \tag{4.33}$$

respectively.

When looking at (4.28) and (4.29) the question arises, how to choose the threshold λ_1 for the single user detection in slot A and with that P_{det}^A and P_{false}^A , which are necessary for evaluating Equations (4.30) to (4.33). Note that the false alarm probability can always be derived from the detection probability with the used detector model. This reduces the question on how to determine P_{det}^A . Looking again at Figure 4.12, the complex interactions of the detectors in different slots and different SUs can be seen. Since all SUs are assumed to be identical and due to the symmetry of slot B and slot C, the system basically contains two concatenated detectors

with the different thresholds λ_1 and λ_2 . One possible approach is to use a fixed P_{dct}^A and thus also a fixed λ_1 which for example can be defined during the system design process. The resulting ROCs after collecting the boosting signals in slot B are depicted in Figure 4.13 for several single user detection probabilities. The mean power of the boosting signal of one SU is related to the noise, the ratio of which is denoted by $\text{SNR}_B = \sigma_B^2 / \sigma_N^2$. The considered P_{dct}^A are the theoretically necessary single user detection probabilities corresponding to the network detection probabilities 0.99, 0.9 and 0.8 according to (4.22) with the given parameters. This fixed approach is for example applied in [81] for deriving the ROCs of the SU-AP while receiving the boosting signals.

Nevertheless, when using a fixed P_{dct}^A , the resulting ROC is actually only optimal for exactly one detection probability P_{dct}^B . The performance of the system can be enhanced when adapting λ_1 as well as λ_2 to minimize the resulting false alarm probability P_{false}^B in slot B. In other words, for each given network detection probability in slot B the choice of λ_1 in slot A influences the shape of the pdf of the detector in slot B, and thus also λ_2 . Algorithm 4.1 describes the necessary steps for generating the resulting ROC when adapting both, λ_1 and λ_2 .

Algorithm 4.1: Generate ROC with dynamically adapted P_{dct}^A .

Input: P_{dct}^B, L
Output: $P_{\text{dct}}^A, P_{\text{false}}^A, P_{\text{false}}^B$

- 1 **foreach** P_{dct}^B **do**
- 2 calculate theoretical P_{dct}^A for given P_{dct}^B and L with (4.22)
- 3 determine corresponding λ_1 with (3.14)
- 4 calculate P_{false}^A with (3.15)
- 5 determine corresponding λ_2 for P_{dct}^B with (4.28) and (4.18)
- 6 calculate P_{false}^B with (4.29) and (4.19)
- 7 **end**

P_{dct}^A is determined by using the given P_{dct}^B as the desired network detection probability in (4.22) and therefore the effects of the boosting phase—for example when not all neighbors are transmitting a boosting signal—are neglected and perfect signaling is assumed. The resulting ROC is also shown in Figure 4.13.

For optimally configured thresholds one would expect that the ROC is always better than the ROCs based on a fixed P_{dct}^A . It can be clearly seen, though, that this is not the case for large P_{dct}^B , implying that the calculated P_{dct}^A is not optimal for the given network detection probability in slot B. This results from neglecting the boosting

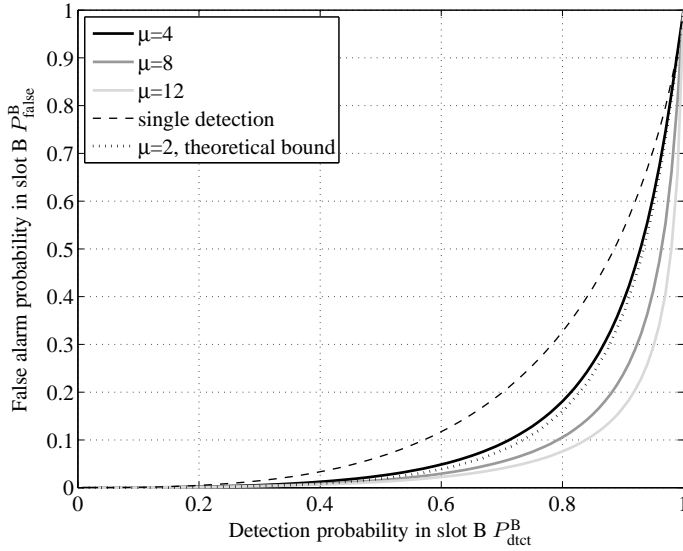


Figure 4.14 ROCs for the detection of the boosting signals in slot B (Mode I) for different μ (SNR = 0 dB, SNR_B = 3 dB and $M = 4$).

effects when determining P_{dct}^A . The optimum can be determined by using (4.28) instead of (4.22) and thus considering the boosting effects for calculating P_{dct}^A when P_{dct}^B is given. However, this turns out to be a complex optimization process, because with (4.18), (4.28) depends on P_{dct}^A as well as on λ_2 . Both parameters now have to be optimized jointly to achieve a minimum P_{false}^B for a given P_{dct}^B . Therefore, Algorithm 4.1 first approximates P_{dct}^A using (4.22), based on which then λ_2 is determined. This approach reduces the involved computation complexity, but only introduces a small degradation of the resulting ROC as can be seen in Figure 4.13. Due to the complex optimization process and the comparatively small performance gain, for the following investigations the ROCs are derived with Algorithm 4.1. This does not degrade the quality of the derived statements.

Figures 4.14 to 4.17 show the corresponding ROCs for varying μ , SNR, SNR_B and M , respectively. The performance of the detection subsystem increases with an increasing number of expected neighbors μ . However, it is of course not as good as the theoretically possible performance, when comparing the results in Figure 4.14 and Figure 4.5. Note that the ROC for $\mu = 4$ in slot B is close to the theoretical ROC for $\mu = 2$ since the probability that slot B is used for boosting is 0.5. A significant performance gain is also achieved especially when increasing the SNR

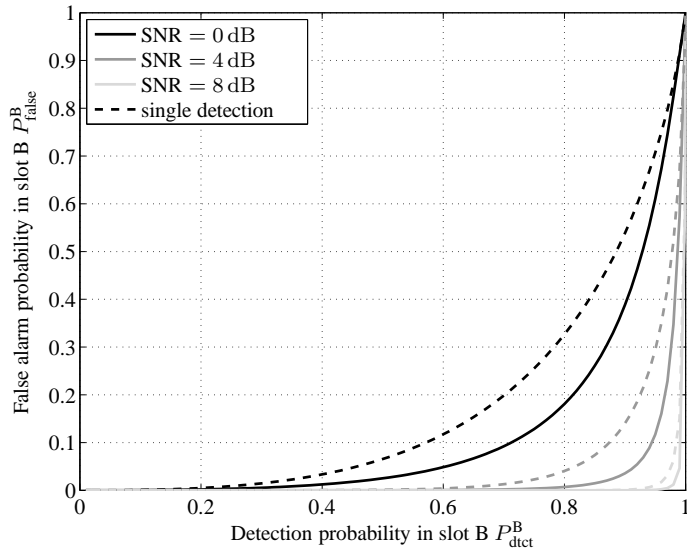


Figure 4.15 ROCs for the detection of the boosting signals in slot B (Mode I) for different SNR ($\mu = 4$, $\text{SNR}_B = 3$ dB and $M = 4$).

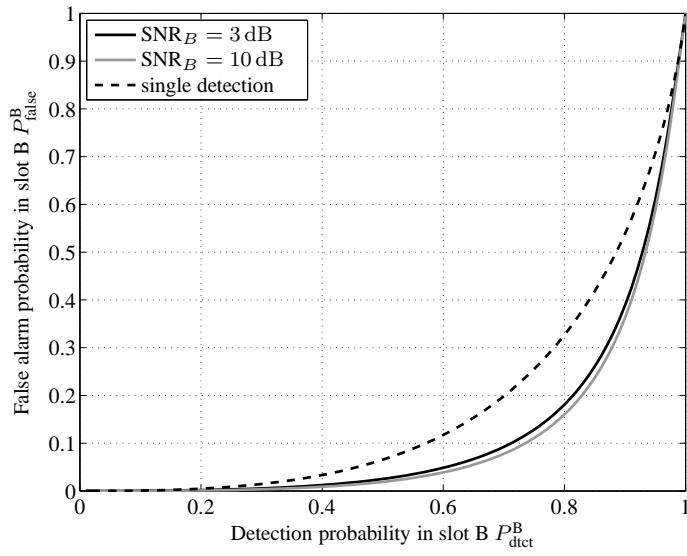


Figure 4.16 ROCs for the detection of the boosting signals in slot B (Mode I) for different SNR_B ($\mu = 4$, $\text{SNR} = 0$ dB and $M = 4$).

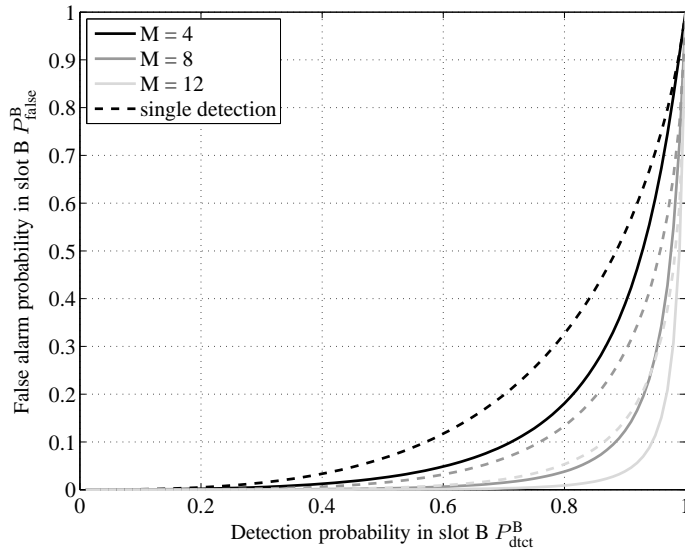


Figure 4.17 ROCs for the detection of the boosting signals in slot B (Mode I) for different M ($\mu = 4$, $\text{SNR} = 0$ dB and $\text{SNR}_B = 3$ dB).

(Figure 4.15) or the number of samples M (Figure 4.17), whereas the SNR_B has only little influence (Figure 4.16).

Detection and False Alarm Probabilities for Slot C

For deriving the detection and false alarm probabilities in slot C for an SU in mode II, again first the transmitted boosting signals of the stations in mode I are considered. Similar to mode II, an SU only sends a boosting signal if it previously detected a PU signal. But in contrast, now there are three possibilities for deciding when to transmit a boosting signal in slot C. The decision either can be based only on the single user detection result P_{dct}^A achieved in slot A (type A):

$$\begin{aligned}
 P_{\text{boost}}^C &= 0.5 \cdot P_{\text{dct}}^A && \text{for PU} \\
 P_{\text{boost}}^C &= 0.5 \cdot P_{\text{false}}^A && \text{for } \overline{\text{PU}},
 \end{aligned} \tag{4.34}$$

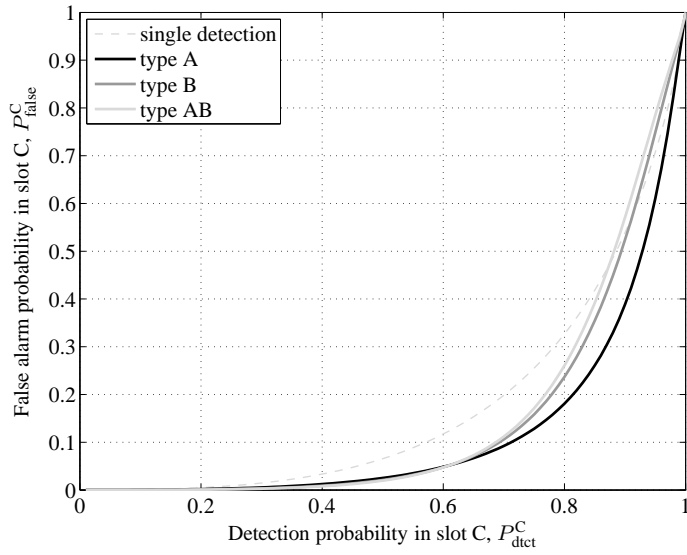


Figure 4.18 Comparison of the ROCs for the detection of the boosting signals in slot C (Mode II) depending on the type of input detection probability.

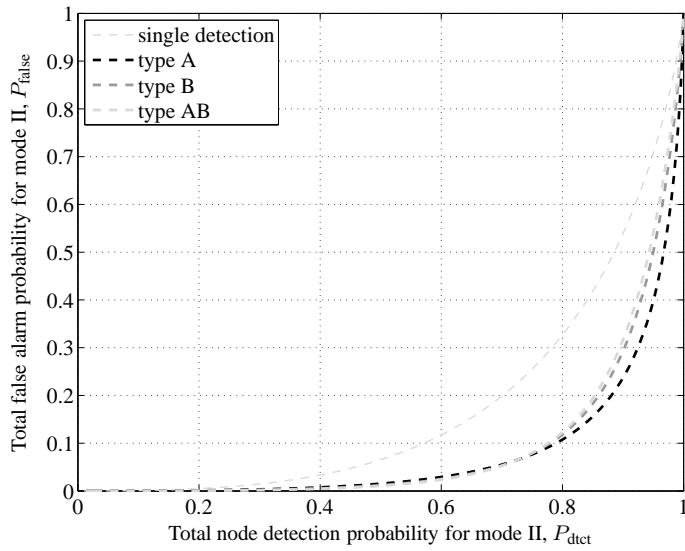


Figure 4.19 Final ROC of a node after completing all three phases depending on the type of input detection probability.

solely on the detection result in the collecting phase (slot B) P_{dct}^B as derived in (4.28) (type B):

$$\begin{aligned} P_{\text{boost}}^C &= 0.5 \cdot P_{\text{dct}}^B && \text{for PU} \\ P_{\text{boost}}^C &= 0.5 \cdot P_{\text{false}}^B && \text{for } \overline{\text{PU}}, \end{aligned} \quad (4.35)$$

or on a combination of both (type AB):

$$\begin{aligned} P_{\text{boost}}^C &= 0.5 \cdot (1 - (1 - P_{\text{dct}}^A) \cdot (1 - P_{\text{dct}}^B)) && \text{for PU} \\ P_{\text{boost}}^C &= 0.5 \cdot (1 - (1 - P_{\text{false}}^A) \cdot (1 - P_{\text{false}}^B)) && \text{for } \overline{\text{PU}} \end{aligned} \quad (4.36)$$

where P_{boost}^C denotes the probability that an SU transmits a boosting signal in slot C and the factor 0.5 represents the assumed probability that slot C is used as a boosting phase. The resulting ROCs are depicted in Figure 4.18. Type A results in an identical ROC as for slot B. The results show that using the detection results of slot B as input for boosting in slot C (either solely or in combination with P_{dct}^B) does not achieve the best performance. The reason is that the single user detection and false alarm probability in slot A are optimized with respect to an optimum performance in slot B, which is identical to type A in slot C. Furthermore, the boosting process not only signals and amplifies detected existing PU signals but also falsely detected signals, therefore also increasing the overall false alarm probability and degrading the system performance.

Figure 4.19 finally shows the corresponding ROCs for the different types after an SU operating in mode II has completed all three phases and combined the detection results from slot A and slot C. Since type A achieves the best performance and at the same time is equivalent to the performance when in mode I, this ROC is valid for all nodes in the network and therefore also represents the average network performance.

Optimization of the Slot Lengths

The results of the previous subsections were based on the assumption that the slots A,B,C are equally sized, that is, each slot contains the same number of complex valued samples M . This does not necessarily result in the best possible performance for a given total number of available samples M_{total} in a detection period. The number of samples allocated to each slot is denoted by M_A , M_B and M_C with the lower index indicating the corresponding slot, where

$$M_{\text{total}} = M_A + M_B + M_C. \quad (4.37)$$

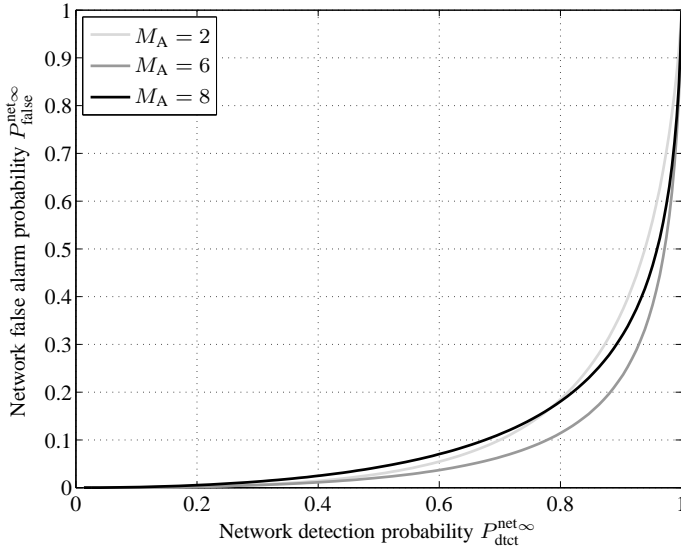


Figure 4.20 ROCs for $M_{\text{total}} = 12$ and different M_A .

Since both, slot B and C, can be either used as a boosting or collection phase, they are assumed to be equal-sized, that is, $M_B = M_C$. Figure 4.20 shows the ROCs for a detection period with $M_{\text{total}} = 12$ and a different number of samples M_A allocated to slot A. The size of slot B (and slot C) is then

$$M_B = M_C = \frac{M_{\text{total}} - M_A}{2}, \quad (4.38)$$

for $M_{\text{total}} - M_A$ yielding an even number. Due to this condition, it is better to vary M_B and then determine the corresponding M_A . The results in Figure 4.20 show that the performance for $M_A = 6$ is better than for $M_A = 2$ and $M_A = 8$. For this reason, now the question is discussed how to optimally allocate the available samples M_{total} to the three slots, resulting in a minimum network false alarm probability. For given M_{total} and M_B , the corresponding network false alarm probability can be determined by using Algorithm 4.1 when using the appropriate values for M . For calculating the single user probabilities P_{dct}^A and P_{false}^A (lines 2–4), M_A has to be used, whereas for calculating P_{dct}^B and P_{false}^B (lines 5 and 6), M_B as given in (4.38) must be used. Figure 4.21 depicts the resulting network false alarm probabilities for $\text{SNR} = 6$ dB, $P_{\text{dct}}^{\text{net}\infty} = 0.999$ and $3 \leq M_{\text{total}} \leq 30$, depending on the number of samples used for boosting, that is, $M_B + M_C$.

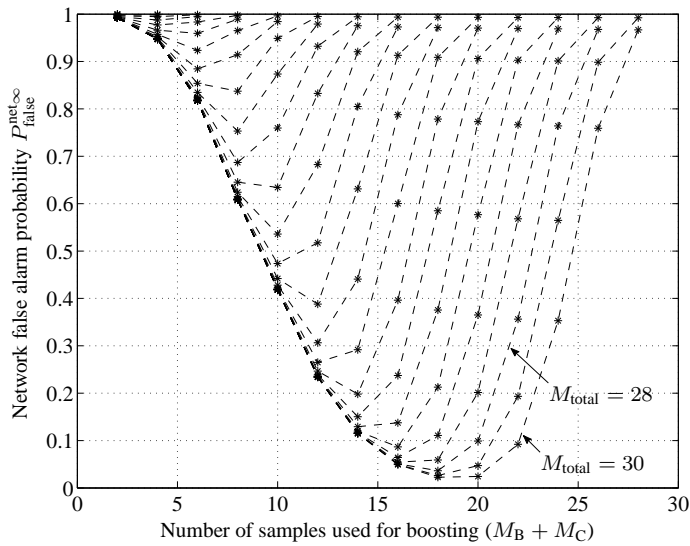


Figure 4.21 False alarm probability depending on the number of samples used for slot B for various M_{total} (SNR = 6 dB).

Especially for large M_{total} , the configuration of the slot lengths has tremendous impact on the performance of the detection subsystem. The achieved false alarm probabilities range from values near 1 which makes the operation of an overlay system impossible, to values close to 0. Furthermore, all curves show a minimum for one specific configuration of the slot lengths which represents the favorable operating point. Note that an analytical solution for determining this operating point is not feasible due to the complexity of the involved calculations. Therefore, the plots were generated numerically. When determining the number of samples used for slots B and C resulting in the optimum operating point and relating it to M_{total} , Figure 4.22 can be derived. It shows the fraction of M_{total} that has to be used for slots B and C to achieve the best performance, depending on M_{total} and for different SNRs. The sawtooth characteristic of the curves results from the fact that M_{total} is increased in steps of 1, but M_B and M_C can only be increased jointly. From Figure 4.22 it can be concluded that in general it is better to use a larger fraction of M_{total} for the single user detection in slot A for smaller SNRs as well as a smaller M_{total} .

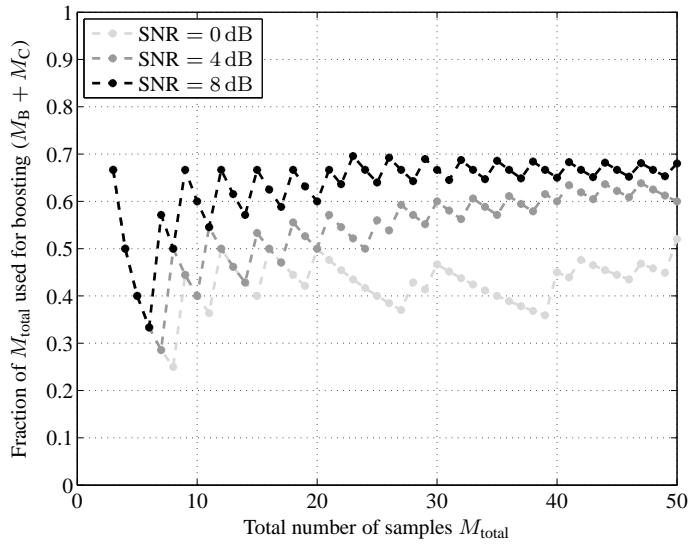


Figure 4.22 Fraction of M_{total} that should be used for slots B/C to achieve a minimum false alarm probability depending on M_{total} .

4.4 Medium Access Control for Overlay Systems in Ad Hoc Mode

The distributed detection approach discussed in the previous sections is based on the assumption that all SUs perform the detection phase at the same time. This ensures that only the PU signal is detected and that the detection result is not affected by signals from the SU system. Therefore, the detection phases must be coordinated within the SU system which is done in the MAC layer according to Section 3.2. In an SU system operating in infrastructure mode the detection phases can be easily coordinated by the SU-AP. It sends a special control frame which includes a defined block for the detection period, so when an SU receives this frame it exactly knows when to be silent.

When considering an SU system operating in ad hoc mode, the coordination of the detection phases becomes more difficult because of the lacking access point and therefore the missing possibility of a centralized coordination. In the following, the resulting challenges are discussed and a new approach for the medium access control for overlay systems in ad hoc mode—abbreviated with AHOMAC—is proposed.

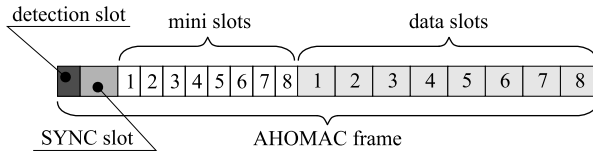


Figure 4.23 AHOMAC frame with mini slots and data slots.

4.4.1 Coordination of Spectrum Occupancy Measurements

An overlay node in ad hoc mode has to meet two main challenges in order to establish communication links to other overlay nodes. As in every overlay scenario it firstly has to consider information about the PU system's spectrum occupation in order to avoid interference. Additionally, due to the ad hoc properties of the overlay system, the node needs information about other overlay nodes existing in the neighborhood. In this section, an approach for the MAC layer is proposed that considers both aspects. Note that neighboring nodes can be located in a different area of the licensed system and thus may detect a different spectrum occupation vector. This leads to the necessity of exchanging spectrum occupation vectors between the overlay nodes which thus may only use the intersection of the available subcarriers for communication, as discussed in Section 3.5.2.

MAC Frame Structure

Many common MAC protocols for ad hoc systems are based on CSMA/CA techniques, as for example the distributed coordination function (DCF). Nevertheless, in the context of an overlay scenario this approach is not suitable due to its sensing property. As stated above, all overlay nodes have to pause their transmissions and perform the measurement at the same time, requiring a periodic structure with defined slots for detection periods.

Since CSMA/CA is an unsuitable technique for this application, the focus lies on another class of MAC protocols: synchronized protocols based on TDMA. In [68] a protocol called multiservices dynamic reservation (MDR) TDMA is proposed with the purpose to serve as a general MAC framework combining constant bit rate (CBR), variable bit rate (VBR) and packet data. Similar to packet reservation multiple access (PRMA) a TDMA frame is divided into two blocks of slots: The first block contains a set of mini slots whereas the second part has the same amount of data slots. For the proposed ad hoc overlay MAC (AHOMAC) protocol [18] the concept of mini slots and data slots is basically adopted. Figure 4.23 shows

the applied general frame structure. Additionally, two special slots are inserted at the beginning of the AHOMAC frame. The detection of the PU system's spectrum occupation is performed in the first slot and is thus used to handle the overlay property of the SU system. The second added slot is called SYNC slot and is used to handle the ad hoc property of the system. Based on the beacons that are transmitted in the SYNC slot, neighboring nodes can be detected and the timing of the AHOMAC frames of the involved SUs is synchronized. The remaining slots are available for data transmissions. In general the number and size of data and mini slots can be arbitrary, depending on the required update interval.

AHOMAC Frame Synchronization and Node Detection

A new overlay node entering the scenario at first has neither information about the available subcarriers nor about other overlay nodes nearby. Thus, the overlay node always starts operating in scanning mode and performs the following steps in order to retrieve the necessary information:

1. In the first step it takes a measurement of the spectrum occupation in order to derive the current spectrum occupation vector. This can only be a first estimate, since the coordination with other overlay nodes has not been performed until then. Furthermore, it listens for beacons from other overlay nodes.
2. In case the node cannot detect any beacons within a defined time, it assumes that there are no other overlay nodes nearby and starts sending beacons itself during the SYNC slot. During times when the node is not transmitting, it listens for beacons sent by other overlay nodes nearby. Only subcarriers in channels that are not occupied by the PU system are used.
3. If the node detects a beacon it adjusts the beginning of its frame according to the offset included in the received beacon and switches to association mode.

After performing these steps, at least two nodes form a synchronized cluster in which spectrum occupation measurements can be taken and data connections can be set up. Since there is no difference between the nodes (for example master/slave assignment), each overlay node has to follow the same set of rules. Thus, each overlay node always tries to send a beacon during the SYNC slot, even if a node already received a beacon in the same SYNC slot. However, the beacon is only sent when no other beacon is detected at the scheduled time to avoid a collision of beacons. The advantage is that new nodes entering the edge of the cluster will recognize the cluster and can synchronize to it, thus increasing its size. Note that

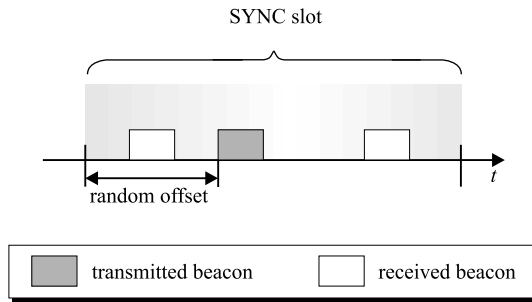


Figure 4.24 SYNC slot with several beacons.

several beacons fit into one SYNC slot as depicted in Figure 4.24 and that each beacon is sent with a random offset within the SYNC slot in order to decrease the probability of collisions. This offset is included in each beacon. From all received beacons the one that was sent earliest is chosen for a frame re-synchronization, all others are only used for the detection of nearby nodes. Figure 4.25 shows the frame synchronization and node detection as a flow chart.

Data Transmission Coordination

In the following the main steps for creating a data connection are described. It is assumed that frame synchronization was already performed successfully, only the available subcarriers are used and the node has knowledge about other overlay nodes within its transmission range. In order to setup a connection to a neighboring node, the initiative node performs following actions, which are similar to reservation TDMA:

1. By observing the mini slots and the corresponding data slots the node recognizes ongoing transmissions in the overlay system and determines available slots, one of which is chosen at random.
2. Similar to CSMA/CA an request to send (RTS) is sent within the chosen data slot, delayed by a random offset. The upper bound of the offset is determined in a way that the RTS always arrives completely within the data slot. During the offset the node is listening whether or not another node already has sent an RTS. Collisions can occur when several nodes want to use the data slot and send a RTS at the same time. In this case the procedure is repeated in the next frame with a different offset. If a node still has not been successful

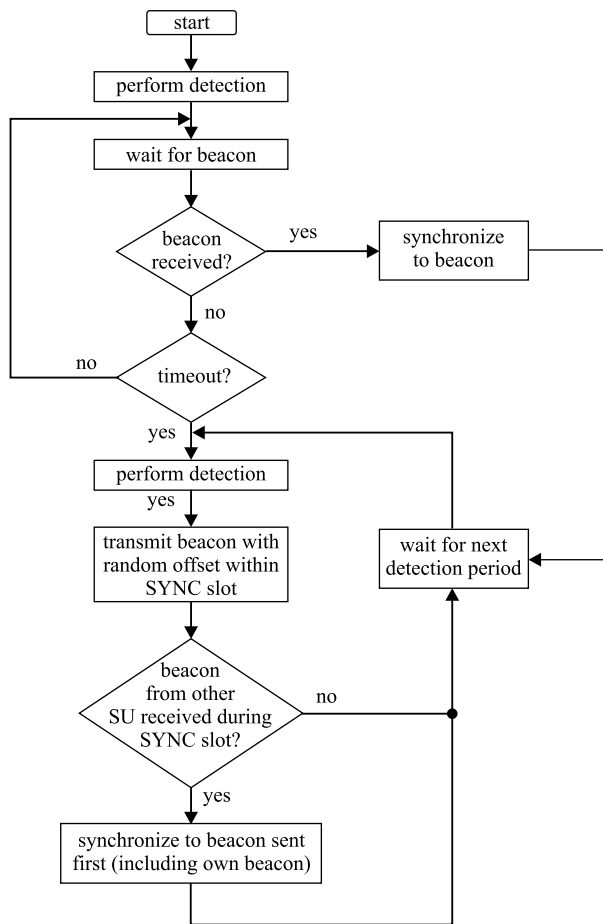


Figure 4.25 Flow chart for the synchronization of the detection periods and node detection.

it chooses a different data slot.

3. The receiver receives the RTS and answers with a clear to send (CTS) in the corresponding mini slot. The connection is now established and by observing the data and mini slots all other nodes within the transmission range of both communicating nodes are notified about the current transmission. Thus, this method avoids hidden nodes. The number of successive frames for the established connection can be limited in order to provide fairness among all participating nodes.

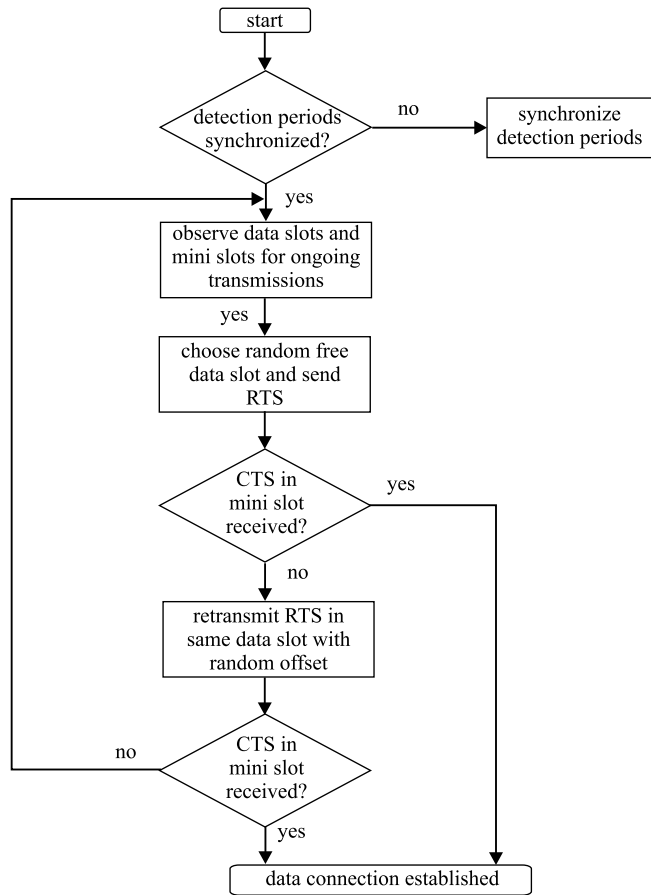


Figure 4.26 Flow chart for setting up a data connection.

5 Detection of Spectral Resources in a Multiband Environment

In this chapter the considered overlay scenario is extended from a single frequency band to multiple frequency bands in which the SU system can operate, as illustrated in Figure 5.1. The bandwidth in which the SU system can operate simultaneously is limited by the available hardware and especially by the analog-digital and digital-analog converters. Although higher bandwidths can be achieved with current technology, often the costs and the resolution constrain their application in mobile devices for the mass market. Therefore, the overall available frequency range is subdivided in several frequency bands. In this context the focus is shifted from the detection of the PU system's current spectrum occupation to the detection of spectral resources. This type of detection becomes important in a multi-band scenario, since the SU system can only operate in one frequency band at the same time.

In contrast to the detection of the current spectrum occupation, the goal of spectral resource detection is to obtain an overview on the estimated occupation situation in all frequency bands in the system. In case of a shortage of resources in the active frequency band, the SU system then can switch to another frequency band with currently more resources available. Since the SU system cannot operate in more than one frequency band at the same time, it has to tune in to all frequency bands periodically to perform a measurement. During this time no transmissions can be

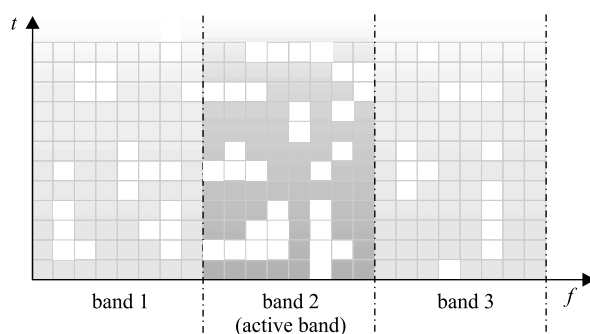


Figure 5.1 Multi-band scenario.

made, forcing the SU system to find an optimal trade-off between spending time for transmitting data and time for increasing the accuracy of information regarding the estimated occupation situation in the other frequency bands by performing detections. So basically, the SU system constantly must decide whether it will transmit data, perform a detection or switch to another frequency band.

Thus, the detection subsystem of an SU system in a multi-band environment has two main objectives. To clearly point out the differences to the previous sections, the main characteristics of both parts are summarized in the following. Note that in contrast to the previous section now it is assumed that the SU system is operating in infrastructure mode and has an access point which coordinates the detection for all SUs and decides when to use which frequency band.

Detection of the PU System's Spectrum Occupation

This part is responsible for the exact and detailed detection of the available resources in the current frequency band. The main characteristics are:

- Detection is performed frequently enough to avoid collisions with the PU system.
- Detection is performed for all subchannels at the same time by using an FFT.
- The detection results are used directly to determine the exact available resources.

Detection of Spectral Resources

The objective of this part of the detection subsystem is to generate a coarse overview on the average available resources in all frequency bands. This information is used to make more global and strategic decisions, in which frequency band an operation would be best in near future. The main characteristics are:

- Detection is performed once in a while to obtain an estimation of the current occupation situation. In contrast to the detection of the PU system's spectrum access, the goal is not to avoid collisions with the PU system.
- Detection is performed using an FFT, but only one frequency band can be observed at a time. When detecting in a different frequency band, the subsequent update interval cannot be used for transmissions in the active/current frequency band.

- The detection results are only used for a mid-term estimation and prediction of the spectrum occupation for making strategic decisions. They have no immediate impact on the ongoing transmission.

This type of problem can be interpreted as a reinforcement learning problem, which is an approach from the area of artificial intelligence. Similar to an animal or human being here an agent (implemented in the detection subsystem of the SU system) learns which decisions are good and which are not so good (or even bad) in different situations by receiving different rewards and feedback depending on the decision made in each situation.

In the following subsections, a short introduction to reinforcement learning is given, including a formal model for representing the general learning problem and solution methods. A detailed introduction including further references can be found in [73]. This is followed by specific problem formulations based on a Markov decision process (MDP) for the detection of spectral resources, depending on different assumptions. Suitable learning strategies are then chosen, modified and finally evaluated in their performance.

Related work includes for example [88], where a partially observable MDP is used to derive a class of decentralized cognitive MAC protocols for opportunistic spectrum access networks. Although also an MDP formulation is used, a different scenario is assumed and the difference between the detection of the PU system's current allocation and the estimation of spectral resources is not considered.

5.1 Reinforcement Learning

Reinforcement learning is an area from machine learning and artificial intelligence which is based on the interdisciplinary concept of learning by interacting with the environment. It transfers the concept of learning to machine domain. Like a human being that learns to achieve a goal by applying a certain behavior, in reinforcement learning an agent wants to achieve a goal and learns which actions lead to success and which to failure by performing them and evaluating the feedback received from the environment. One important difference to other learning approaches is that reinforcement learning is not a supervised learning method (as for example applied in optical character recognition where a training set of known characters is required). This means that there is no teacher instructing the agent in advance which action is good or bad. The agent has to find out on its own which possible action—or sequence of actions—leads to a situation closer to the long-term goal. In other

words, the agent tries to influence the environment by applying a certain behavior, observes the response of the environment and adapts its behavior if necessary. In consequence, the model of the environment does not necessarily have to be known by the agent which is another advantage compared to supervised learning.

Reinforcement learning can be seen as a framework using the elements *policy*, *reward function* and *value function* for representing the artificial intelligence problem. The policy defines the behavior of an agent. The reward function indicates the immediate reward for an action and is defined by the environment, whereas the value function defines what is good from a long-term perspective. Modern reinforcement learning dates back to the late 1980s, but has its roots in two independent disciplines which then came together. The first one is optimal control with the concept of value functions and the solution methods using dynamic programming [40]. Nevertheless, the model of the environment has to be known and computations quickly become very complex. The second discipline is the concept of learning by trial-and-error that started in the psychology of animal learning and then led to the first approaches of artificial intelligence. The relevant trial-and-error learning approaches for reinforcement learning are assumed to be selectional as well as associative. That is, the selection of alternatives is done based on a comparison of their consequences and are additionally associated with the particular situations. Nevertheless, the main focus of research in the area of artificial intelligence soon shifted to supervised learning concepts such as pattern recognition and perceptual learning. Another area of artificial intelligence are evolutionary methods. They are distinguished from reinforcement learning by the lacking concept of value functions. Evolutionary methods perform a search directly in the policy space.

When facing a learning problem that has to be solved by reinforcement learning, the formulation of the problem within the reinforcement learning framework plays an important role. Then, depending on the available information regarding the environment, usually an existing reinforcement learning method can be used, or adapted if necessary. Therefore, first the framework for a formal problem formulation is discussed, followed by the three main learning approaches of reinforcement learning. Although dynamic programming is actually not a learning algorithm, it provides the theoretical foundation for reinforcement learning. It can be interpreted as a method for searching an optimal policy and can be applied if the model is known and the problem is not too complex. In contrast, Monte Carlo methods do not depend on a model of the environment. They estimate the value functions by interaction with the environment, that is, on-line or by simulation, but improvements are only made on an episode-by-episode basis. Finally, temporal-difference

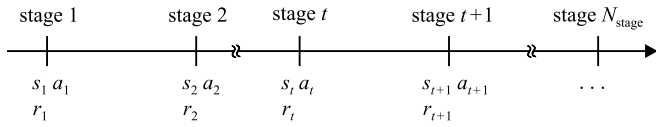


Figure 5.2 States, actions and rewards at each stage.

learning combines the advantages of dynamic programming as well as Monte Carlo methods and is considered as the central and novel idea of reinforcement learning.

5.1.1 Formal Model of the Learning Problem

The reinforcement learning model consists of the agent representing the instance that learns and acts, and the environment as an instance responding to the actions with rewards and presenting new situations to the agent. The environment is defined as everything outside of the agent's absolute control (not outside of its knowledge) and also can be changing dynamically. The interaction between the agent and the environment is assumed to take place at discrete time steps referred to as stages (see Figure 5.2) and is event driven, that is, the time steps are triggered by an event (the action). This kind of system is also referred to as *discrete event system* [32]. Note that the time intervals between two state transitions do not necessarily have to be equidistant.

The agent and the environment are connected by two different types of interaction as shown in Figure 5.3. The agent is equipped with appropriate sensors and is able to observe the current state of the environment. After performing an action in stage t , $t \in \mathbb{N}$, the state of the environment evolves and the agent receives an update about the environment's state s_{t+1} as well as an immediate reward r_{t+1} , indicating how good or bad it was to perform the chosen action a_t in the previous stage. Based on this update information, the cycle begins again and the agent chooses the next action a_{t+1} resulting in the next stage with the new state s_{t+2} and the reward r_{t+2} .

Therefore, the reinforcement learning problem is based on a sequential decision model which can be described by the tuple $\langle \mathcal{T}, \mathcal{S}, \mathcal{A}, p, r \rangle$ [66], where:

- $\mathcal{T} = \{1, 2, \dots, N_{\text{stage}}\}$, $N_{\text{stage}} \leq \infty$ is the set of stages, which can either be finite or infinite. Elements of \mathcal{T} are denoted by t . For a finite N_{stage} the decision model is referred to as a *finite horizon* problem, otherwise it is called an *infinite horizon* problem.

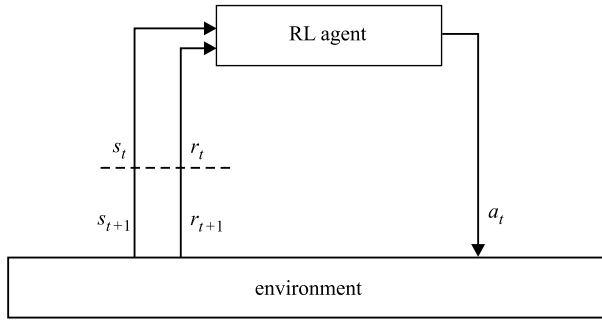


Figure 5.3 General interaction between the agent and the environment.

- \mathcal{S} is the finite set of states the agent can be in.
- \mathcal{A} is the finite set of all actions the agent can perform. Since in general the agent might not be able to perform every action in every state, \mathcal{A}_s describes the set of possible actions in state s , with $\mathcal{A}_s \subseteq \mathcal{A}$ and $s \in \mathcal{S}$. Thus, $\mathcal{A} = \cup_{s \in \mathcal{S}} \mathcal{A}_s$.
- $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0,1]$ defines the state transition probability function $p(s' | s, a)$, giving the probability of transitioning to state $s' \in \mathcal{S}$ after performing the action $a \in \mathcal{A}$ while in state $s \in \mathcal{S}$, with $\sum_{s'} p(s' | s, a) = 1$.
- $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ defines the reward function $r(a, s)$, giving the reward the agent receives when performing action $a \in \mathcal{A}$ while in state $s \in \mathcal{S}$.

After observing the current state, the agent needs to choose an action for the next stage. This is done according to the policy $\pi : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$, where $\pi_s(a)$ defines the probability that action a is executed when the agent is in state s . Note that $\sum_a \pi_s(a) = 1$ for each state s and therefore $\pi_s(a)$ is not a conditional probability. If the Markov property holds, that is, the only state that is relevant for the future behavior is the current state,

$$\begin{aligned} \Pr \{s_{t+1} = s', r_{t+1} = r \mid s_t, r_t, a_t, s_{t-1}, a_{t-1}, \dots, r_1, s_0, a_0\} \\ = \Pr \{s_{t+1} = s', r_{t+1} = r \mid s_t, r_t\}, \quad (5.1) \end{aligned}$$

this corresponds to an MDP. In general, reinforcement learning problems do not necessarily have to be modeled as an MDP to be solvable, but the Markov property

holds for most of them and is also assumed in the following work. Furthermore, in the general case infinite sets of states and actions are possible, but are not considered here.

Learning Goal and Return

The learning goal of the agent is to maximize the expected total reward, also called return. It is a function of the received rewards over time. Among the possible actions the agent chooses the action a_t which results in a maximized return. Therefore, it is relevant whether a finite or infinite horizon is considered. Tasks with an infinite horizon are also called continuing tasks. But there are also tasks (for example when learning how to play a game) which can be split up in several episodes, that is, when the agent arrives at a terminal state the system is reset and a new episode begins. Accordingly, these tasks are referred to as episodic tasks.

For defining the return, episodic and continuing tasks have to be distinguished. For an episodic task, the return can be simply defined as the cumulated received immediate rewards of one episode. When considering a continuing task, the cumulated reward can easily grow infinite. Therefore, in this case a discount factor $\gamma \in [0,1)$ is introduced, discounting future rewards. Defining the expected return as [73]

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}(a_{t+k}, s_{t+k}) \quad (5.2)$$

handles both types of tasks. For episodic tasks an additional absorbing state is introduced in which the agent transits after arriving at the terminal state of the episode. It stays there forever without receiving further rewards. Setting $\gamma = 1$ then cumulates all immediate rewards of an episode from stage t on. With $\gamma \in [0,1)$ the discounted expected return in stage t of a continuing task is considered. The closer γ is to one, the more influence the future rewards have and the agent is more farsighted. When $\gamma = 0$, the future rewards have no impact at all and the agent only tries to maximize the immediate reward. This behavior is also called myopic.

Value Functions

The concept of value functions plays a fundamental role in reinforcement learning. A value function maps a real value to a state or an action, indicating how good it is to be in that state or to perform that action, in terms of the expected return. This of course also depends on the policy the agent is following. Therefore, the value of a state is defined as the expected return when starting in state s and thereafter

following the policy π :

$$V^\pi(s) = \mathbb{E}_\pi \{R_t \mid s_t = s\} = \mathbb{E}_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s \right\}. \quad (5.3)$$

In a similar way, the value of performing action a when in state s and thereafter following the policy π is defined as

$$Q^\pi(s, a) = \mathbb{E}_\pi \{R_t \mid s_t = s, a_t = a\} = \mathbb{E}_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right\}. \quad (5.4)$$

$V^\pi : \mathcal{S} \rightarrow \mathbb{R}$ and $Q^\pi : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ are called state-value function and action-value function for policy π , respectively. The objective of solving a reinforcement learning problem is to find an optimal policy, that is, a policy that maximizes the return in a long-term and therefore also the corresponding optimal value-functions. The optimal state-value function is denoted as

$$V^*(s) = \max_{\pi} V^\pi(s), \quad \forall s \in \mathcal{S} \quad (5.5)$$

and the optimal action-value function as

$$Q^*(s, a) = \max_{\pi} Q^\pi(s, a), \quad \forall s \in \mathcal{S} \quad \text{and} \quad \forall a \in \mathcal{A}. \quad (5.6)$$

Generalized Policy Iteration

For finding the optimal policy π^* and the optimal state-value function $V^*(s)$, reinforcement learning algorithms follow the idea of generalized policy iteration [73]. A policy is first evaluated, that is, the value function is calculated or estimated. In the second step, the policy is improved with respect to the state-value function. The improved policy is then again evaluated. These two interacting steps are iteratively repeated as illustrated in Figure 5.4 until no changes are required any more, resulting in the optimal value function and the optimal policy. Evaluation and improvement of the policy are the two core components of reinforcement learning algorithms and they occur in dynamic programming approaches, Monte Carlo methods as well as temporal-difference learning algorithms as summarized in the following.

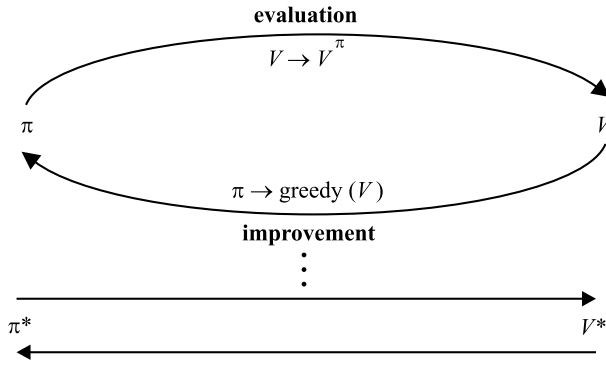


Figure 5.4 Interaction of value and policy functions in the generalized policy iteration.

5.1.2 Dynamic Programming

Dynamic programming algorithms use value functions to search for good policies for problems formulated as an MDP. They require a given perfect model of the environment and quickly result in great computation effort. Therefore, dynamic programming mainly serves as a theoretical foundation for other reinforcement learning algorithms, which basically try to relax these limitations. For a better understanding in this section deterministic policies are assumed, but the results can be applied to stochastic policies as well.

Bellmann Equations

Based on known state transition probabilities and expected rewards, the state-value function $V^\pi(s)$ (5.3) can also be written in the following way, which is also known as the corresponding Bellmann equation:

$$V^\pi(s) = \sum_a \pi_s(a) \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V^\pi(s')] \quad (5.7)$$

with the state transition probabilities $\mathcal{P}_{ss'}^a$

$$\mathcal{P}_{ss'}^a = \Pr \{s_{t+1} = s' \mid s_t = s, a_t = a\} \quad (5.8)$$

and the expected rewards

$$\mathcal{R}_{ss'}^a = \mathbb{E} \{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\}. \quad (5.9)$$

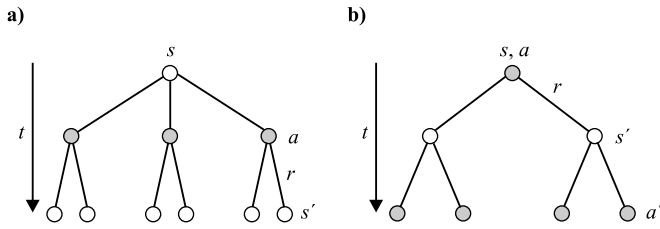


Figure 5.5 Dependencies of state, action, reward and successor state for (a) $V^\pi(s)$ and (b) $Q^\pi(s, a)$.

The Bellmann equation states that the value of a state depends on the value of the successor state and the corresponding received reward. All possible succeeding states with the rewards are weighted by the probabilities of occurring, that is, by the policy $\pi_s(a)$ and the state transition probability $\mathcal{P}_{ss'}^a$. Figure 5.5 illustrates the dependencies of state, action, reward and successor state for $V^\pi(s)$ and $Q^\pi(s, a)$, where time is evolving from the top to the bottom. Note that the value of a state depends on all possible succeeding states. Estimates for a state thus depend on estimates for other states. This is called *bootstrapping*.

The corresponding Bellmann optimality equations for V^* and Q^* can be then written as

$$\begin{aligned} V^*(s) &= \max_a \mathbb{E} \{ r_{t+1} + \gamma V^*(s_{t+1}) \mid s_t = s, a_t = a \} \\ &= \max_a \sum_{s'} \mathcal{P}_{ss'}^a \{ \mathcal{R}_{ss'}^a + \gamma V^*(s') \} \end{aligned} \quad (5.10)$$

and

$$\begin{aligned} Q^*(s, a) &= \mathbb{E} \left\{ r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') \mid s_t = s, a_t = a \right\} \\ &= \sum_{s'} \mathcal{P}_{ss'}^a \left[\mathcal{R}_{ss'}^a + \gamma \max_{a'} Q^*(s', a') \right], \end{aligned} \quad (5.11)$$

stating that the value of a state under an optimal policy is equal to the expected return when choosing the best action in that state. Note that the Bellmann optimality equations do not depend on a specific policy.

Policy Evaluation

With the assumption of a completely known model, (5.7) results in a system of $|\mathcal{S}|$ equations with $|\mathcal{S}|$ unknowns and the value function $V^\pi(s)$ can in principle be

computed. Nevertheless, the values usually have to be estimated due to the complexity of the computations. One suitable approach is an iterative approximation method where the Bellmann equation is used as an update rule:

$$\begin{aligned} V_{k+1}(s) &= \mathbb{E}_\pi \{r_{t+1} + \gamma V_k(s_{t+1}) \mid s_t = s\} \\ &= \sum_a \pi_s(a) \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V_k(s')] \end{aligned} \quad (5.12)$$

with $k \in \mathbb{N}$ denoting the iteration. It can be shown that the sequence $\{V_k\}$ converges to V^π for $k \rightarrow \infty$. The corresponding algorithm for iterative policy evaluation is given in Algorithm 5.1 [73].

Algorithm 5.1: Policy Evaluation.

Input: the policy π to be evaluated and an arbitrarily initialized $V(s)$

Output: $V \approx V^\pi$

```

1 repeat
2    $\Delta \leftarrow 0$ 
3   foreach  $s \in \mathcal{S}$  do
4      $v \leftarrow V(s)$ 
5      $V(s) \leftarrow \sum_a \pi(s, a) \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V(s')]$ 
6      $\Delta \leftarrow \max(\Delta, |v - V(s)|)$ 
7   end
8 until  $\Delta < \theta$  (a small positive number)
```

Policy Improvement

The value function of a policy can now be used for improving the policy, based on the policy improvement theorem [73]. It states that if for all $s \in \mathcal{S}$

$$Q^\pi(s, \pi'(s)) \geq V^\pi \quad (5.13)$$

holds for two deterministic policies π and π' , then π' must be as good as, or better than, π . In other words, if π' is at least as good as π it must result in an equal or greater expected return

$$V^{\pi'}(s) \geq V^\pi(s) \quad (5.14)$$

for all states $s \in \mathcal{S}$. A short proof of this theorem is given in [73]. So basically, choosing an action $\pi'(s) = a \neq \pi(s)$ that results in a greater value than the current

policy leads to an improvement. This suggests applying a greedy policy

$$\pi'(s) = \arg \max_a Q^\pi(s, a) = \arg \max_a \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V^\pi(s')] \quad (5.15)$$

which chooses the action with the greatest value. Furthermore, it can be shown that if the new policy is not better than the old one, both policies are optimal policies [73]. Algorithm 5.2 [73] shows the corresponding policy improvement algorithm, where *policy-stable* indicates whether the policy could be improved or not.

Algorithm 5.2: Policy Improvement.

Input: $V(s), \pi(s)$
Output: an improved policy $\pi(s)$; *policy-stable*

- 1 *policy-stable* \leftarrow true
- 2 **foreach** $s \in \mathcal{S}$ **do**
- 3 $b \leftarrow \pi(s)$
- 4 $\pi(s) \leftarrow \arg \max_a \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V(s')]$
- 5 **if** $b \neq \pi(s)$ **then** *policy-stable* \leftarrow false
- 6 **end**

Policy Iteration

In the previous sections methods were presented how to evaluate a given policy and how to improve a policy based on its evaluation. Both methods can be applied iteratively to obtain the optimal value function and the corresponding optimal policy as shown in Algorithm 5.3 [73]. Note that the input for the policy evaluation is the improved policy resulting from the policy improvement step. The iterations are continued until the policy improvement does not result in an improved policy anymore.

Algorithm 5.3: Policy Iteration.

- 1 Initialize $V(s) \in \mathbb{R}$ and $\pi(s) \in \mathcal{A}_s$ arbitrarily for all $s \in \mathcal{S}$
- 2 *policy-stable* \leftarrow false
- 3 **repeat**
- 4 evaluate $\pi(s)$ with Algorithm 5.1
- 5 improve $\pi(s)$ with Algorithm 5.2
- 6 **until** *policy-stable*

Value Iteration

When looking at policy iteration the question comes up, whether the iteration in the policy evaluation part has to be fully completed until the state values converged, which can result in a large computation cost. In fact, it can be shown that even one single iteration can be performed without losing overall convergence of the policy iteration algorithm. This approach is called value iteration and can be expressed as

$$\begin{aligned} V_{k+1}(s) &= \max_a \mathbb{E} \{ r_{t+1} + \gamma V_k(s_{t+1}) \mid s_t = s, a_t = a \} \\ &= \max_a \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V_k(s')], \end{aligned} \quad (5.16)$$

for all $s \in \mathcal{S}$. Value iteration combines policy evaluation and policy improvement in a single step and can also be interpreted as the Bellmann optimality function (5.10) turned into an update rule. Algorithm 5.4 [73] describes the value iteration algorithm. It terminates when the value function does not change significantly anymore.

Algorithm 5.4: Value Iteration.

Input: discount factor γ
Output: an optimal policy $\pi(s)$

- 1 initialize $V(s)$ arbitrarily for all $s \in \mathcal{S}$
- 2 **repeat**
- 3 $\Delta = 0$
- 4 **foreach** $s \in \mathcal{S}$ **do**
- 5 $v \leftarrow V(s)$
- 6 $V(s) \leftarrow \max_a \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V(s')]$
- 7 $\Delta \leftarrow \max(\Delta, |v - V(s)|)$
- 8 **end**
- 9 **until** $\Delta < \theta$ (*a small positive number*)
- 10 Output a deterministic policy π , such that
- 11 $\pi(s) \leftarrow \arg \max_a \sum_{s'} \mathcal{P}_{ss'}^a [\mathcal{R}_{ss'}^a + \gamma V(s')]$

5.1.3 Monte Carlo Methods

In contrast to dynamic programming, Monte Carlo methods do not require complete knowledge of the system dynamics. Nevertheless, they have the same objective, that is, finding an optimal policy with the help of value functions. In both

cases, this is done according to the concept of the generalized policy iteration. Due to the missing model in the Monte Carlo methods, this cannot be achieved by solving a system of equations. Instead, the value functions are estimated on-line, that is, while interacting with the environment, or by simulation. Note that it is often possible to simulate the interaction with the environment although the complete system dynamics cannot be described by an explicit model, that is, the rewards and state transition probabilities are not known. Therefore, Monte Carlo methods can be considered as learning algorithms (in contrast to dynamic programming).

Value functions are estimated by averaging samples of complete returns and therefore, only episodic tasks are considered. Updates of the value function and policy improvements are only done from episode to episode. Furthermore, in case of lacking knowledge of the model, it is not sufficient to know the state-values. It is also necessary to estimate the action-values to be able to find a policy. For estimating the value of an action $Q^\pi(s, a)$, the agent starts in state s , performs action a and thereafter follows the policy π . This is repeated several times and the resulting returns of an episode are averaged.

Another important difference to dynamic programming is that Monte Carlo methods do not bootstrap. That is, the estimates for each state are independent of other states. Furthermore, the values of a single state can be estimated with a computational effort that is independent of the total number of states. However, for determining an optimal policy, it is required that all state–action pairs are visited once in a while. To improve a policy, also the alternative actions not included in the current policy have to be evaluated to be able to choose the best action. A tradeoff between exploring new state–action pairs and exploiting the already obtained information has to be made.

Exploitation Versus Exploration

A very simple method for ensuring that every state–action pair is visited is called *exploring starts*. For each episode the state–action pair to start with is chosen randomly and with a non-zero probability. This way, each pair will be visited an infinite number of times in the long run assuming an infinite number of episodes. Nevertheless, this is not in all scenarios possible.

Another approach is the concept of ε -soft policies. A policy is defined as an ε -soft policy, if $\pi_s(a) \geq \frac{\varepsilon}{|\mathcal{A}_s|}$ holds for all states and actions, for $0 < \varepsilon < 1$. A special case of ε -soft policies are ε -greedy policies, which choose with probability $1 - \varepsilon + \frac{\varepsilon}{|\mathcal{A}_s|}$ the greedy action, that is, the action with the highest value (exploitation). Accordingly, with probability $\frac{\varepsilon}{|\mathcal{A}_s|}$ one of the remaining actions is chosen

(exploration). Note that $\frac{\varepsilon}{|\mathcal{A}_s|} \geq 1$.

In contrast to ε -greedy policies, where in the exploration case all actions are selected with an equal probability, the softmax action selection method assigns a higher probability to those actions with a greater value. Thus, the estimates of actions with low values still are being updated, but not as frequently as actions with a higher value. The probabilities for choosing an action a are for example given by a Boltzmann distribution

$$\Pr(a) = \frac{e^{Q(s,a)/\xi}}{\sum_b e^{Q(s,b)/\xi}}, \quad (5.17)$$

where ξ can be used to adjust the behavior. For $\xi \rightarrow 0$ actions with the highest Q -value are preferred, whereas for large ξ all actions are chosen with nearly the same probability, approaching the behavior of the ε -greedy action selection method.

On-Policy and Off-Policy Monte Carlo Control

In Section 5.1.2 policy improvement was discussed for deterministic policies. Based on the policy improvement theorem it can also be shown that policy improvement works for ε -soft policies. This is necessary when working with action-values, due to the required trade-off between exploitation and exploration. On the way to the optimal action-value function, any ε -greedy policy is an improvement over any ε -soft policy. So when looking at the general policy iteration again, instead of making the policy greedy during the policy improvement phase, it is only moved toward a greedy policy by making it ε -greedy, and therefore still allowing exploration.

This approach is called an *on-policy* method because the policy that is evaluated and improved is also used to choose the actions and to generate the episodes. Algorithm 5.5 [73] shows an on-policy, ε -soft Monte Carlo control algorithm. In contrast, *off-policy* methods use separate policies. The behavior policy controls the actions and has to ensure that all state–action pairs are visited once in a while, that is, it assumed to be ε -soft. The estimation policy is the policy that is improved and can also be deterministic. Note that if the probability of a state–action pair is nonzero in the estimation policy, it also is required to be nonzero in the behavior policy.

5.1.4 Temporal-Difference Learning

Temporal-difference learning is the central idea of reinforcement learning and combines the advantages of dynamic programming and Monte Carlo methods. Re-

Algorithm 5.5: On-policy, ε -soft Monte Carlo control algorithm.

```

1 Initialize, for all  $s \in \mathcal{S}$ ,  $a \in \mathcal{A}_s$ :
2    $Q(s, a) \leftarrow$  arbitrary
3    $Returns(s, a) \leftarrow$  empty list
4    $\pi \leftarrow$  an arbitrary  $\varepsilon$ -soft policy
5 repeat forever
6   Generate an episode using  $\pi$ 
7   foreach pair  $s, a$  appearing in the episode do
8      $R \leftarrow$  return following the first occurrence of  $s, a$ 
9     Append  $R$  to  $Returns(s, a)$ 
10     $Q(s, a) \leftarrow$  average( $Returns(s, a)$ )
11 end
12 foreach  $s$  in the episode do
13    $a^* \leftarrow \arg \max_a Q(s, a)$ 
14   forall  $a \in \mathcal{A}_s$  do
15      $\pi(s, a) \leftarrow \begin{cases} 1 - \varepsilon + \frac{\varepsilon}{|\mathcal{A}_s|} & \text{if } a = a^* \\ \frac{\varepsilon}{|\mathcal{A}_s|} & \text{if } a \neq a^* \end{cases}$ 
16   end
17 end

```

call that dynamic programming methods bootstrap, that is, estimates for a state are based in part on estimates of other states, and value updates as well as policy improvements can be done incrementally in every stage. However, Monte Carlo methods learn by interacting with the environment and therefore it is not required to know the complete model. These advantages are unified in temporal-difference (TD) learning methods allowing an update in every stage by interacting with the environment without knowing the system dynamics as model. Temporal-difference learning also follows the generalized policy improvement and consists of a policy evaluation part (value estimation) and a policy improvement part. The simplest form for the estimation of state-values in temporal-difference learning is TD(0), using

$$V(s_t) \leftarrow V(s_t) + \alpha [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \quad (5.18)$$

as update rule for the estimates of state-values. TD(0) is a special case of the popular TD(λ) which is also used in TD-Gammon, a program for playing backgammon [74, 75] and an impressive application of reinforcement learning. With only

Algorithm 5.6: TD(0) for estimating V^π .

```

1 Initialize  $V(s)$  arbitrarily,  $\pi$  to the policy to be evaluated
2 repeat for each episode
3   Initialize  $s$ 
4   foreach step of episode do
5      $a \leftarrow$  action given by  $\pi$  for  $s$ 
6     Take action  $a$ ; observe reward  $r$  and next state  $s'$ 
7      $V(s_t) \leftarrow V(s_t) + \alpha [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]$ 
8      $s \leftarrow s'$ 
9   end
10 until  $s$  is terminal

```

little backgammon knowledge it learned to play on the level of the world's best human players. Nevertheless, in this work only TD(0) is considered, because it allows a comprehensive discussion of the relevant aspects when applying TD-learning to the detection of spectral resources. However, it is expected that a better performance can be achieved when using more sophisticated TD-learning methods, but the investigated basic principles and mechanisms remain unchanged regardless of the specific TD-learning method. The complete algorithm for TD(0)-learning is given in Algorithm 5.6.

A general rule for incrementally updating estimates is given as

$$\text{NewEstimate} = \text{OldEstimate} + \text{StepSize}[\text{Target} - \text{OldEstimate}].$$

With each update, the estimate is moved towards a desirable value, the *target*. The expression $[\text{Target} - \text{OldEstimate}]$ represents the error, that is, how far the current estimate is away from the target. The step-size parameter α indicates the influence of the new information, that is, how fast the estimate is moved towards the target. The target can be noisy, but nevertheless gives the direction in which to move the estimation.

Basically, Monte Carlo methods and temporal-difference learning both use this general update rule, but they are distinct in the applied target. Constant- α Monte Carlo for example applies the following update rule

$$V(s_t) \leftarrow V(s_t) + \alpha [R_t - V(s_t)] \quad (5.19)$$

and uses the return R_t after one episode as target, whereas in TD(0) according to (5.18) the target is set to $r_{t+1} + \gamma V(s_{t+1})$. Temporal-difference learning does

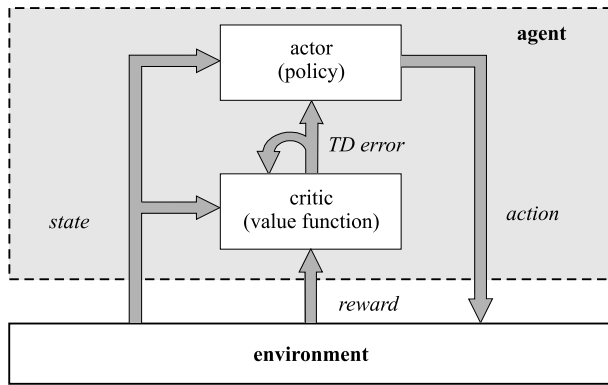


Figure 5.6 Interaction of the secondary user with the environment and the applied actor-critic structure.

not wait until an episode is completed to update the estimates. Instead, the target is composed of the received immediate reward r_{t+1} and the discounted estimate of the succeeding state. Recall from Section 5.1.2 that γ is the discount factor, representing the influence of future states on the estimate. Therefore, temporal-difference learning is also a bootstrapping method. Until now only policy evaluation was considered, and therefore in the following two subsections policy iteration algorithms using state-values and action-values are discussed, respectively.

Actor-Critic Methods

Actor-critic methods are a suitable approach for implementing the general policy iteration for TD methods when using state-values. Figure 5.6 shows the structure of the actor-critic approach, visualizing one of the main advantages of these methods. The state-values and policy are stored in independent memory structures. The actor is used for choosing the actions and thus represents the policy structure. The actions are evaluated by the critic, which therefore represents the state-value structure. To evaluate the action, the critic determines the TD-error

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t), \quad (5.20)$$

which is the second part of (5.18). Equation (5.18) is then also used by the critic to update the state-values. The TD-error is additionally signaled to the actor which maintains preference values for all state-action pairs. They are updated based on

the TD-error according to

$$p(s_t, a_t) \leftarrow p(s_t, a_t) + \beta \delta_t, \quad (5.21)$$

where β is another step-size factor. Finally, the actor uses the preference values for choosing the next action, for example by applying a soft-max action selection method:

$$\pi_{s,t}(a) = \Pr \{a_t = a \mid s_t = s\} = \frac{e^{p(s,a)}}{\sum_b e^{p(s,b)}} \quad (5.22)$$

Note that actor-critic methods are always on-policy, since the critic evaluates the actor which controls the behavior of the agent.

Q-learning

Q-learning is a TD-learning algorithm based on action-values. Similar to Monte Carlo methods, there are on-policy and off-policy TD algorithms, depending whether the behavior policy is equivalent to the estimation policy or not. An example for an on-policy TD-learning algorithm is state-action-reward-state-action (SARSA) (originally called modified Q-learning [71]), but more important for reinforcement learning is Q-learning, which was formulated by Watkins in 1989 [79]. It is an off-policy learning algorithm and therefore a proof for the convergence can be easily performed [80], which makes this algorithm popular. Note that proving the convergence of learning algorithms is one of the major challenges in this field. The update rule for Q-learning is given by

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]. \quad (5.23)$$

The optimal action-value function is approximated by using the action with the greatest action-value and the immediate reward as target. The complete Q-learning algorithm is described by Algorithm 5.7 [73].

5.2 Problem Formulation as Markov Decision Process

In this section the previously introduced reinforcement learning methods are applied to the spectral resource detection problem. Since the SU system does not know in advance how many resources are available in each frequency band, it has to find a good trade-off between exploitation and exploration, that is, between playing

Algorithm 5.7: Q -learning algorithm

```

1 Initialize  $Q(s, a)$  arbitrarily
2 repeat for each episode
3   Initialize  $s$ 
4   foreach step of episode do
5     Choose  $a$  from  $s$  using policy derived from  $Q$  (e. g.,  $\epsilon$ -greedy)
6     Take action  $a$ , observe  $r, s'$ 
7      $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$ 
8      $s \leftarrow s'$ 
9   end
10 until  $s$  is terminal

```

safe and staying in the current frequency band where it knows how much data it can send, and switching to other frequency bands with the chance of finding one with more resources, but also with the risk of switching to a band with less resources. In the following, three different models with increasing complexity regarding the assumptions are developed and corresponding solution strategies proposed. The first model completely neglects signaling aspects and thus can be modeled as an n -armed bandit problem serving as a reference model. The second model includes signaling aspects and is formulated as an MDP. Finally, the third model additionally distinguishes the costs for switching to another frequency band and the costs for performing an out-of-band detection.

5.2.1 Reference Model

To begin with, it is assumed that switching the SU system to another frequency band does not result in additional costs, so the SU system can choose an arbitrary frequency band at each stage and immediately transmit data. When using this simple assumption for the spectral resource detection problem, it can be formulated as a learning problem similar to the n -armed bandit problem [73]. It has only a single state but several actions $a_n \in \mathcal{A}$ with $n = 1, \dots, N_{\text{fb}}$. Each action corresponds to the cycle of choosing a frequency band n , performing one detection phase and transmitting data [17]. Thus, the time between two stages corresponds to the length of an update interval. The received immediate reward is the amount of data that was transmitted during the corresponding update interval. Note that the reward directly depends on the number of currently available subcarriers, which is determined by the detection phase at the beginning of each period. The action-values $Q^*(a_n)$

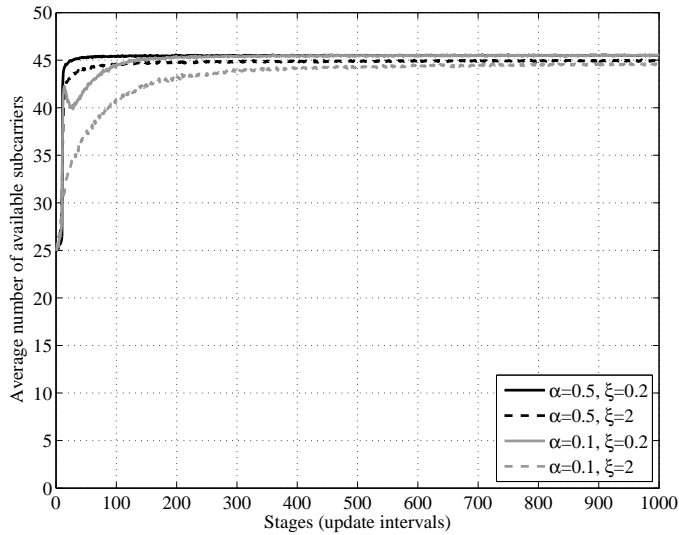


Figure 5.7 n -armed bandit formulation: average learning curves for different α and ξ .

represent the expected reward when performing a_n .

The goal is to find the frequency band in which the SU system can currently transmit most data, that is, the action with the highest value $Q^*(a_n)$. Since the SU system does not know the true Q -values it has to estimate them. This can be done by averaging the received rewards over time, assuming a stationary process for the spectrum occupation of the PU system in each frequency band. The estimated Q -values can be updated incrementally by applying the following update rule every time the action was executed, which is based on the general update rule discussed in Section 5.1.4:

$$Q_{k+1}(a_n) = Q_k(a_n) + \alpha[r_{k+1} - Q_k(a_n)] \quad (5.24)$$

where α is a step-size parameter and r the received reward. Note that the index k is not identical to the current stage t , but counts the number of times a specific action was performed. For choosing the actions the softmax action-selection method (5.17) introduced in Section 5.1.3 can be applied. It gives a trade-off between exploration and exploitation based on the estimated Q -values. Figure 5.7 shows the averaged learning curves for a scenario with $N_{fb} = 10$ frequency bands containing $N_{sc} = 50$ channels for different combinations of α and ξ . Recall from

Section 5.1.3 that for large ξ in (5.17) all frequency bands are chosen with nearly the same probability. Each channel of the PU system is assumed to be available with the probability $P_{\text{avail}}(n)$ with n denoting the frequency band, resulting in different binomial distributions with the parameters N_{fb} and $P_{\text{avail}}(n)$ for the number of available subcarriers in each frequency band. Note that P_{avail} is the same for each channel within one frequency band. All simulated learning curves show the average over 4000 episodes and for each episode and each frequency band $P_{\text{avail}}(n)$ was sampled from a uniform distribution which then stayed constant for the entire episode and therefore simulating a stationary scenario. The results shown in Figure 5.7 illustrate the basic characteristics of the n -armed bandit problem. When using a larger learning rate α the curve converges faster to its optimum. Furthermore, the influence of ξ on the softmax action selection method can be observed. For large ξ smaller Q -values are chosen more often leading to more time spent with exploration and therefore the long-term performance is not as good as for small ξ . However, these curves show a higher performance in the short term. Exploration is an advantage in the beginning but when the good values are found it causes a performance degradation. The local optimum for $\alpha = 0.1$ and $\xi = 0.2$ results from the start values for the Q -values which were set to the maximum number of possibly available channels, that is, to N_{sc} . This encourages exploration even if the frequency band with the most available channels was already visited. Due to its simplicity, this scenario serves as a reference for the following scenarios.

5.2.2 Basic Model

In a more realistic scenario, switching the frequency band is not involved with zero costs, since the complete SU system with all participating stations has to be informed about the scheduled changes. This process requires additional signaling and time, and therefore reduces available resources for transmitting data. In contrast to the reference model with only one state and multiple actions, now a model with several states is used, based on the MDP formulation introduced in Section 5.1.1. The achieved reward in each stage depends on the action as well as the current state of the environment.

System Model and Signaling Issues

The SU system is operating in infrastructure mode where the access point is responsible for the decision which frequency band to use. Therefore, the learning agent is embedded into the SU-AP where it is placed in the MAC layer part of the detection subsystem as depicted in Figure 5.8.

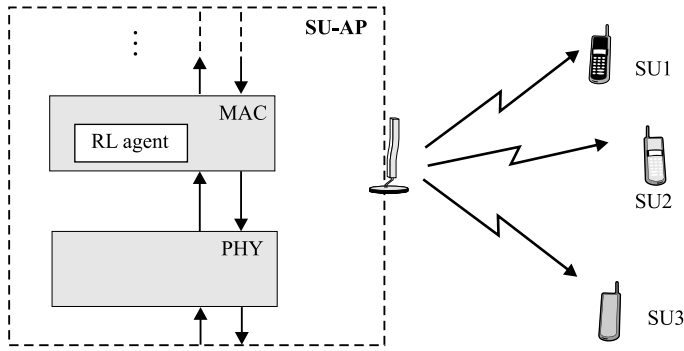


Figure 5.8 Infrastructure scenario and location of the reinforcement learning (RL) agent in the secondary user access point (SU-AP).

Note that the RL agent is not identical with the SU system or the SU-AP. Moreover, in this case even part of the SU system belongs to the environment which can be observed by the RL agent as pointed out below. All other SUs are synchronized to the SU-AP regarding the behavior of their detection subsystem, that is, all SUs always operate in the same frequency band and the timing of their detection phases is synchronized. The signaling of the detection results is done according to the boosting protocol discussed in Section 4.3 and is completed within one update interval.

The SU-AP initiates a transition to another frequency band by sending a switching control packet. The associated SUs send an acknowledgement (ACK) and then switch to the indicated frequency band. The SU-AP stays in the original frequency band and waits until all ACKs of the associated SUs are received before switching to the new frequency band. If necessary, the SU-AP retransmits the switching control packet. When changing to another frequency band the SUs stay idle until they receive a broadcast packet from the SU-AP indicating that the SU system has successfully switched to the new band and data transmissions can be continued. The resulting signaling overhead mainly depends on the number of necessary ACK-messages, and therefore on the number of associated SUs. For the following considerations the total amount of signaling overhead is assumed to be constant for each frequency band transition and is denoted by C_{sig} . The number of available resources in each frequency band is modeled as a stochastic process with the random variables $U_n(t)$ at stage t ($n = 1, \dots, N_{\text{fb}}$) and the corresponding samples $u_n(t)$.

States

The set of states the RL agent can be in is defined as

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \setminus \{(n, \text{Sig}_n)\} \quad \text{with } n = 1, \dots, N_{\text{fb}}. \quad (5.25)$$

Since the SU system can only be in one frequency band at a time, the first component of the state

$$\mathcal{S}_1 = \{1, \dots, N_{\text{fb}}\} \quad (5.26)$$

indicates the current frequency band. Depending on the available resources in the current frequency band and the amount of signaling overhead for switching to another frequency band, it is possible that not all signaling information can be transmitted within a single update interval, that is, the SU system stays in signaling mode for more than one stage. Therefore, the second component of the state indicates the current mode of the SU system

$$\mathcal{S}_2 = \{\text{Tx}, \text{Sig}_1, \dots, \text{Sig}_{N_{\text{fb}}}\}, \quad (5.27)$$

where Tx denotes the transmission mode and Sig_n stands for the signaling mode with the intention of switching to frequency band n . Since it is not reasonable to switch to frequency band n while it already is the current frequency band, the states (n, Sig_n) are excluded from the set of possible states. In the following, transmission states are denoted by $\mathcal{S}_{\text{tx}} = \{(n, \text{Tx})\}$ and signaling states by $\mathcal{S}_{\text{sig}} = \{(n, \text{Sig}_j)\}_{n \neq j}$ ($n = 1, \dots, N_{\text{fb}}$ and $j = 1, \dots, N_{\text{fb}}$). The resulting total number of possible states is $|\mathcal{S}| = N_{\text{fb}}^2$.

By introducing the second state component it is possible to consider situations where signaling takes longer than one update interval and still let one stage correspond to one update interval. This simplifies the definition of the rewards and the mapping from available subcarriers to a data rate. A different possible approach would be to define the state as the current frequency band and define a state transition after all signaling is completed, regardless of the necessary time (that is, using stages that are not equidistant). In this case, though, it would be more difficult to describe the resulting reward (cost) when signaling takes longer than a single update interval. Note that switching frequency bands usually becomes necessary when the current resources run low, that is, only a small number of subcarriers is available. This increases the probability that the transmission of the signaling data is distributed over several update intervals.

Actions

The actions that the RL agent can execute represent the decisions the SU system has to make regarding the choice of the frequency band. In each frequency band, the SU system can either transmit data in the current band or switch transmission to another frequency band which comes along with signaling overhead, and thus, decreasing the available resources for data transmissions. Therefore, in the basic model the set of possible actions in a transmission state $s \in \mathcal{S}_{\text{tx}}$ is $\mathcal{A}_{\text{tx}} = \{a_{\text{tx}}, a_{\text{sw} \rightarrow \hat{s}}\}_{s \neq \hat{s}}$ with $\hat{s} \in \mathcal{S}$ and $|\mathcal{A}_{\text{tx}}| = N_{\text{fb}}$. In a signaling state $s \in \mathcal{S}_{\text{sig}}$ there is only one possible action available for the SU system. It has to continue to transmit the signaling information until signaling is completed. Therefore the set of possible actions is given by $\mathcal{A}_{\text{sig}} = \{a_{\text{sig}}\}$. The overall set of actions is then given by $\mathcal{A} = \mathcal{A}_{\text{tx}} \cup \mathcal{A}_{\text{sig}}$. The actions are described as follows:

- a_{tx} : perform a detection phase in the current frequency band s and transmit data in the available subcarriers.
- $a_{\text{sw} \rightarrow \hat{s}}$: switch the SU system to frequency band \hat{s} . This action also contains a detection phase in the current frequency band, but instead of data the signaling information is transmitted. In case that there are more resources available than needed for signaling they are used for transmitting data.
- a_{sig} : continue to transmit signaling information.

Figure 5.9 shows the transition graph for the basic model with $N_{\text{fb}} = 3$ available frequency bands.

State Transition Probabilities and Rewards

Since the exact number of available subcarriers for every stage is unknown, also the state transition probabilities and rewards are only partially known. $p(s' | a, s)$ with $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}$ is the probability that the next state is s' when performing action a in the current state s and $r(s' | a, s)$ is the corresponding achieved immediate reward. Table 5.2.2 gives an overview on the state transition probabilities and rewards in a general form, which is also applicable to the example with $N_{\text{fb}} = 3$ depicted in Figure 5.9. The only clearly known state transition probability is $p((n, \text{Tx}) | a_{\text{tx}}, (n, \text{Tx})) = 1$, stating that the SU system will stay in the same frequency band after transmitting data. In a similar way, it is also clear that no data is transmitted as long as there is still further signaling data waiting for transmission, that is, $r((n, \text{Sig}_j) | a_{\text{sw} \rightarrow j}, (n, \text{Tx})) = 0$ and $r((n, \text{Sig}_j) | a_{\text{sig}}, (n, \text{Sig}_j)) = 0$ (with $n = 1, \dots, N_{\text{fb}}$ and $j = 1, \dots, N_{\text{fb}}$).

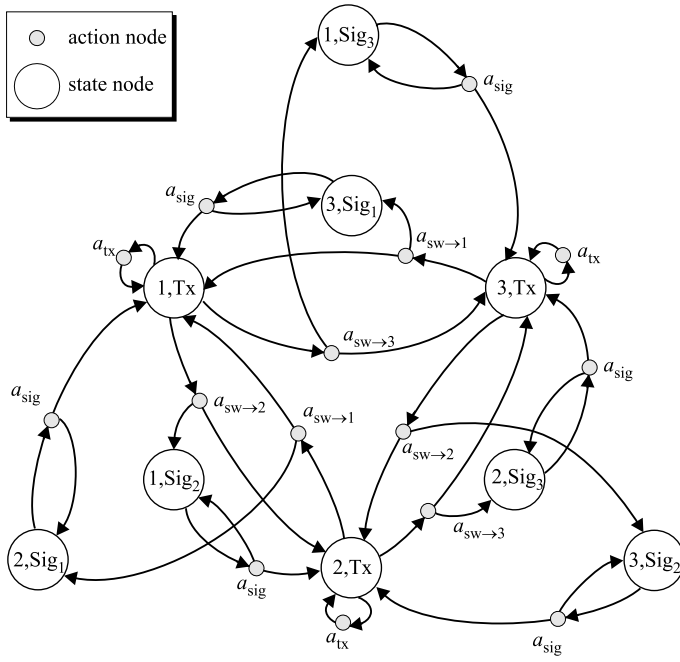


Figure 5.9 Transition graph for the basic model with $N_{\text{fb}} = 3$ available frequency bands.

When initiating a frequency band transition in stage t , two situations can occur. With probability P_{tx} the transmission of the signaling information is completed within the current update interval and the SU system transitions to the target transmission state. $P_{\text{tx}} = p(C_{\text{sig}} \leq u_n(t))$ depends on the size of the signaling overhead and the available resources in the current frequency band and stage. In this case the resulting reward is the number of remaining resources in the stage after transmitting all signaling information, that is, $r((n, \text{Tx}) | a_{\text{sw} \rightarrow j}, (j, \text{Tx})) = u_n(t) - C_{\text{sig}}$. With probability $1 - P_{\text{tx}} = p(C_{\text{sig}} > u_n(t))$ the signaling sequence is not completed within the current update interval and the SU system transitions to the corresponding signaling state while achieving zero reward. The same two situations can occur when the SU system is in a signaling state at stage $t + m$ with $m \in \mathbb{N}$, that is, m states after the frequency band transition was initiated in stage t . Let $c_{\text{sig}}(t + m)$ denote the corresponding remaining signaling overhead where $c_{\text{sig}}(t + 1) = C_{\text{sig}}(t) - u_n(t)$ and $c_{\text{sig}}(t + m + 1) = c_{\text{sig}}(t + m) - u_n(t + m)$.

s	s'	a	$p(s' a, s)$	$r(s' a, s)$
(n, Tx)	(n, Tx)	a_{tx}	1	u_n
(n, Tx)	(j, Tx)	$a_{\text{sw} \rightarrow j}$	P_{tx}	$u_n - C_{\text{sig}}$
(n, Tx)	(n, Sig_j)	$a_{\text{sw} \rightarrow j}$	$1 - P_{\text{tx}}$	0
(n, Sig_j)	(j, Tx)	a_{sig}	P_{tx}	$u_n - c_{\text{sig}}$
(n, Sig_j)	(n, Sig_j)	a_{sig}	$1 - P_{\text{tx}}$	0
----- additional action for enhanced model: -----				
(n, Tx)	(n, Tx)	$a_{\text{dt} \rightarrow \hat{s}}$	1	0

with $n = 1, \dots, N_{\text{fb}}; j = 1, \dots, N_{\text{fb}}$ and $n \neq j$

Table 5.1 Transition probabilities and rewards for the basic model using a general notation.

Either the transmission of the remaining signaling information $c_{\text{sig}}(t + m)$ is now completed (with probability $P_{\text{tx}} = p(c_{\text{sig}} \leq u_n(t + m))$ and the immediate reward $r((n, \text{Sig}_j) | a_{\text{sw} \rightarrow j}, (j, \text{Tx})) = u_n(t + m) - c_{\text{sig}}(t + m)$), or there is even more signaling information to be transmitted (with probability $1 - P_{\text{tx}} = p(c_{\text{sig}} > u_n(t + m))$). In the latter case the SU system stays in the signaling state and receives zero reward.

5.2.3 Learning Strategy for Basic Model

To develop a learning strategy for the basic model it is important to consider the main properties of the spectral resource detection problem. As discussed in the previous section, the state transition probabilities and rewards are only partially known, which excludes a solely dynamic programming approach. Additionally, the scenario is non-episodic, that is, it continues infinitely and does not have a final state in case that the average allocation in the different frequency bands is time variant. This recommends an on-line approach where learning is done while facing the real problem and not in a separate training phase.

Therefore, TD learning is one of the reinforcement learning approaches that is suitable for this type of problem. In the following, the learning strategy for the basic model is investigated, which is based on a state-value function and applies the actor-critic method discussed in Section 5.1.4. This model is basically adopted

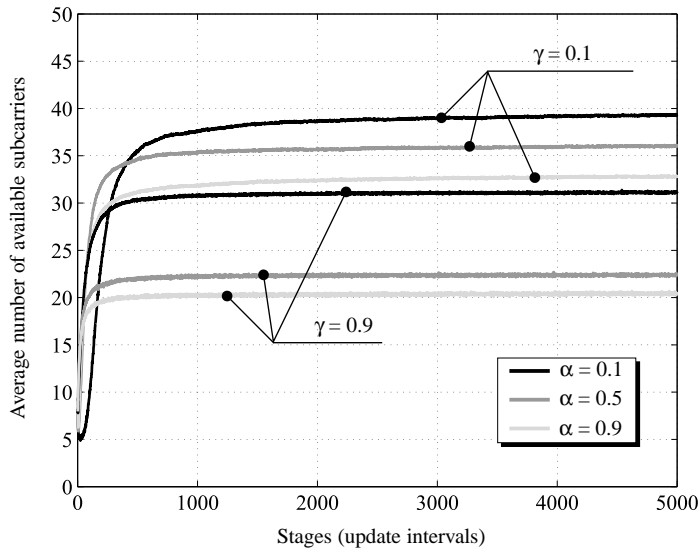


Figure 5.10 Average number of received rewards (available subcarriers) depending on the stage for the basic model using an actor–critic method.

without changes, that is, Equations (5.18) and (5.20) to (5.22) are directly applied to the basic model. One advantage of the actor–critic approach is its easy integration into the cross-layer optimization framework proposed in [43], because it uses a memory structure that explicitly represents the policy independent of the value function and therefore the policy in one layer can be evaluated by a critic in a different layer.

Simulation Results

To evaluate the performance of the learning strategy in combination with the basic model the average performance was investigated by simulations. Similar to the reference model a scenario with $N_{\text{fb}} = 10$ independent frequency bands each with a maximum of 50 available subcarriers was used. For each frequency band the fixed but arbitrary probability $P_{\text{avail}}(n)$ with $n = 1, \dots, N_{\text{fb}}$, giving the probability that a channel within a frequency band is occupied by the PU system, is chosen randomly from a uniform distribution in $(0, 1)$ for each episode. This results in a binomial distribution with the parameters $P_{\text{avail}}(n)$ and N_{fb} for the number of available subcarriers (and thus reward) in each stage and frequency band.

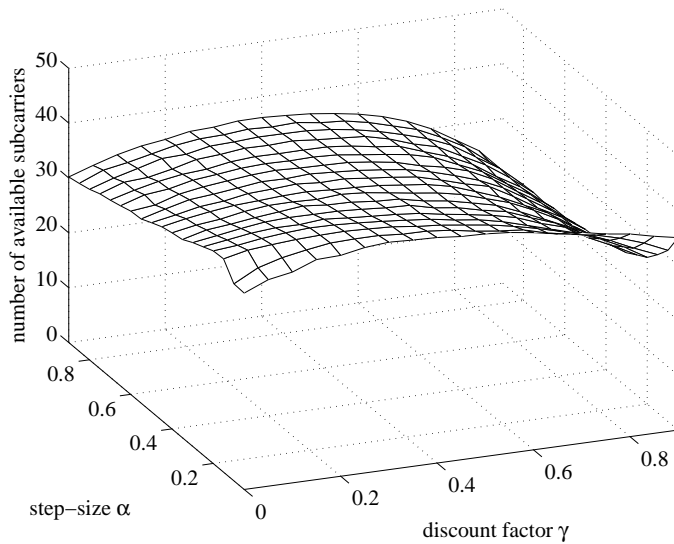


Figure 5.11 Average number of received awards after 10000 stages for a varying step-size α and discount factor γ .

Figure 5.10 shows the learning curves for the step-size factors $\alpha = 0.1$, $\alpha = 0.5$, $\alpha = 0.9$ in Equation (5.18) in combination with the discount factors $\gamma = 0.1$ and $\gamma = 0.9$. For each curve 4000 episodes with 5000 stages were simulated and the average number of achieved rewards in each stage determined, resulting in the displayed curves. The state-values $V(s)$ for the states were initialized to the maximum number of available subcarriers N_{fb} to encourage exploration in the beginning of an episode. Similar to the reference model a small step-size factor achieves a better performance in the long run, but in average it takes longer to find the frequency bands with more available resources. The effect of the discount-factor γ in a long-term is visualized in Figure 5.11. The learning curves for several combinations of α and γ were simulated and the respective average rewards achieved after 10000 stages plotted. When looking at the plot, it can be seen that for a sufficient large α the performance increases with an increasing γ until $\gamma \approx 0.5$. Then the performance decreases again. In other words, considering expected future rewards in the decision making process can be a benefit, but weighting the expected rewards in the far future too strong is disadvantageous. This results from the bootstrapping property of the applied TD-learning algorithm and can be explained when looking

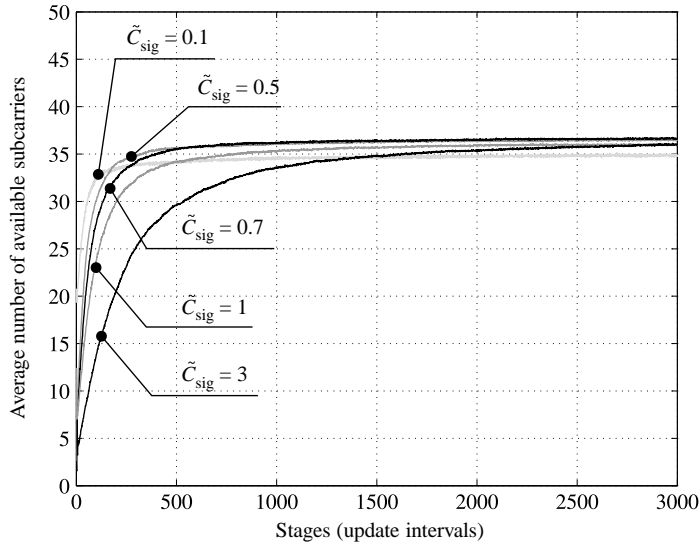


Figure 5.12 Average number of received rewards depending on the stage for different signaling costs \tilde{C}_{sig} (with $\alpha = 0.5$ and $\gamma = 0.5$).

at Equation 5.18. Large γ give a higher weight to the next state value, which is an estimate itself. Only for very small step-size factors even high γ are attenuated enough so that the influence of the future estimates is kept low. Furthermore, Figure 5.11 again illustrates that the long-term performance is better for small step-size factors.

Besides the learning rate, also the signaling overhead has an impact on the performance. Figure 5.12 shows the learning curves for different signaling efforts for $\alpha = 0.5$ and $\gamma = 0.5$. \tilde{C}_{sig} is the signaling overhead related to the maximum resources in one stage if all channels are available to the SU system. This means, that in this case for $\tilde{C}_{\text{sig}} = 1$ signaling takes one stage, but if not all channels are available it takes several stages. The larger the signaling overhead, the longer it takes to find the frequency bands with the most resources, since the SU system spends more time with signaling where no rewards are achieved.

5.2.4 Extended Model with Out-of-Band Detection

Compared with the reference model, the performance of the basic model is not very good. Therefore, an extended model using additional out-of-band detection phases is introduced which enables faster learning and keeping track of the coarse allocation in other frequency bands.

The set of states is the same as in the basic model, whereas the set of actions is extended by an additional action for the transmitting state, resulting in the new set of actions $\mathcal{A}_{\text{Tx}} = \{a_{\text{Tx}}, a_{\text{sw} \rightarrow \hat{s}}, a_{\text{dt} \rightarrow \hat{s}}\}_{s \neq \hat{s}}$ for the enhanced model with $\hat{s} \in \mathcal{S}$ and additionally to the actions listed on page 104

- $a_{\text{dt} \rightarrow \hat{s}}$: perform an out-of-band detection phase in frequency band \hat{s} .

The transition probability is $p((n, \text{Tx}) \mid a_{\text{dt} \rightarrow \hat{s}}, (n, \text{Tx})) = 1$ with the corresponding immediate reward $p((n, \text{Tx}) \mid a_{\text{dt} \rightarrow \hat{s}}, (n, \text{Tx})) = 0$ as summarized in Table 5.2.2, that is, when performing an out-of-band detection the SU system does not receive a reward and stays in the current frequency band for further data transmissions.

Note that performing a detection phase in another frequency band is associated with different costs compared to completely switching to another band. Since the SU system cannot simultaneously perform a detection phase in the current frequency band, it has no reliable information regarding its spectrum occupation. It therefore cannot ensure a transmission without tolerable interference and thus has to omit the transmission of data in the current stage. The out-of-band detection results are transmitted in the target frequency band according to the boosting protocol, exactly like the regular in-band detection results. In contrast to switching to another frequency band, for the out-of-band detection it is not important that it is performed by all SUs so that an acknowledgment can be omitted. If an SU does not receive the detection control packet correctly, it does not participate in the distributed out-of-band detection. In the worst case, it performs a detection phase in a different frequency band and its results do not contribute to the system-wide detection result in the SU-AP.

Figure 5.13 shows an example for a sequence of states and actions with $N_{\text{fb}} = 3$ frequency bands. To begin with, the SU system operates in the second frequency band. At stages 2 and 4 out-of-band detections are performed in frequency band 1 and 3, respectively. In this example, two update intervals are necessary for the signaling of the frequency band transition, which is initiated at stage 5. From stage 7 on, the SU system is transmitting data in frequency band 1.

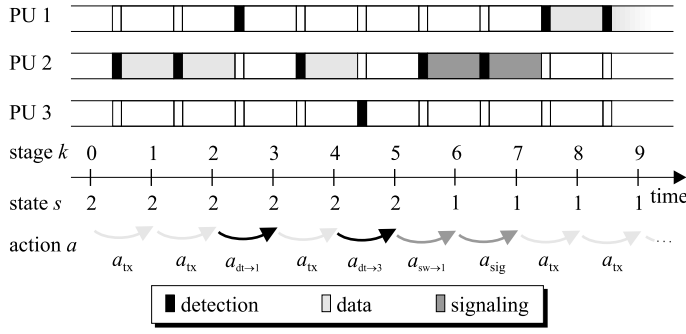


Figure 5.13 Example for state transitions and actions of a SU system operating in a 3-band environment.

5.2.5 Modified Learning Strategy for the Extended Model

When applying the same learning strategy as for the basic model, it is obvious that no performance improvement can be expected, since the additional action never achieves an immediate reward. On the contrary, it even leaves available resources unused. Nevertheless, by performing out-of-band detections, additional information is gained that can be used for improving the decision-making strategy and general performance. In the following a modification of the learning strategy for the extended model is presented and the differences in the actor and critic part of the RL agent are discussed.

Critic

The critic maintains a vector $\hat{\mathbf{u}}(t) = [\hat{u}_1(t), \dots, \hat{u}_{N_{\text{fb}}}(t)]^T$ indicating the estimates of the average number of available resources of each frequency band at stage t . Since in every stage a detection phase is performed in one of the frequency bands, $\hat{\mathbf{u}}$ can be always updated in every stage according to

$$\hat{u}_n(t+1) = \begin{cases} \hat{u}_n(t) + \alpha_2 [u_j(t+1) - \hat{u}_n(t)] & n = j \\ \hat{u}_n(t) & n \neq j \end{cases} \quad (5.28)$$

for $n = 1, \dots, N_{\text{fb}}$, where α_2 is a step-size parameter. The frequency band in which the detection was performed is denoted by j with $j \in \{1, \dots, N_{\text{fb}}\}$. The reliability of the resource estimation vector $\hat{\mathbf{u}}(t)$ is expressed by a corresponding reliability vector $\boldsymbol{\varsigma}(t) = [\varsigma_1(t), \dots, \varsigma_{N_{\text{fb}}}(t)]^T$ with $0 \leq \varsigma \leq 1$. Small ς indicate a low reliability and large ς a high reliability of the respective resource estimation. To

update the reliability values, the critic needs to know which action was performed. In each stage, only the reliability value of the frequency band is increased in which the detection phase was performed. All other reliability values are decreased. Note that the received reward r_{t+1} is not relevant for updating the reliability vector. One possible update rule for ς is

$$\varsigma(t+1) = \varsigma(t) + \alpha_3 [\mathbf{d}(t) - \varsigma(t)] \quad (5.29)$$

where $\mathbf{d}(t) = [d_1(t), \dots, d_{N_{\text{fb}}}(t)]^T$ with $\mathbf{d}(t) \in \{0,1\}^{N_{\text{fb}}}$ is a binary vector, indicating for each frequency band whether a detection phase was performed in the current stage ($d = 1$) or not ($d = 0$). $\alpha_3 \in (0,1)$ is another positive step-size parameter adjusting the influence of new values. Note that every element of ς is updated in every stage.

Now the reliability values have to be combined with the estimated values of the available resources. This is done by

$$\Upsilon_n(t) = \varsigma_n(t) \left(\hat{u}_n(t) - \frac{N_{\text{fb}}}{2} \right) + \frac{N_{\text{fb}}}{2} \quad (5.30)$$

for each $n = 1, \dots, N_{\text{fb}}$. Based on the estimated resources and the reliability for each frequency band a value $\Upsilon_n(t)$ is determined, indicating how good it is to choose the frequency band for data transmissions. This function implements the following properties:

- Reliable information (that is, $\varsigma_n = 1$) about a frequency band n with many available resources is of high value.
- Reliable information about a frequency band n with little available resources is of low value.
- Unreliable information ($\varsigma_n = 0$) about any frequency band has medium value, since there is a chance that there are many resources available, but also the chance of only little resources.

The critic updates the state-values according to the following update rule:

$$V(s_t) \leftarrow V(s_t) + \beta [r_{t+1} + \gamma \Upsilon(s_{t+1}) - V(s_t)], \quad (5.31)$$

that is, compared to the learning strategy for the basic model (c. f. Equation (5.18)), the target of the update rule is now based on Υ . Therefore, also the TD-error is modified:

$$\delta_t = r_{t+1} + \gamma \Upsilon(s_{t+1}) - V(s_t). \quad (5.32)$$

Actor

Based on the TD-error δ and the reliability values ς the actor now has to update the current policy. This is done by calculating preference values p for each action, based on which the policy $\pi_{s,t}(a)$ can then be derived for example by applying the softmax action selection method:

$$\pi_t(s, a) = P(a_t = a | s_t = s) = \frac{e^{p(s,a)}}{\sum_b e^{p(s,b)}}. \quad (5.33)$$

The preferences are calculated in different ways, depending on the type of action:

- Action a_{tx} (transmitting data) is updated using a common update rule:

$$p(s, a_{\text{tx}}) \leftarrow p(s, a_{\text{tx}}) + \beta_1 \delta_t. \quad (5.34)$$

- Actions $a_{\text{sw} \rightarrow \hat{s}}$ switch the SU system to another frequency band. Here it is desirable to switch to a frequency band with a lot of resources if the information about the the resources is reliable. Unreliable information still is preferable in contrast to a certain information about low resources. The following equation gives the mapping:

$$p(s, a_{\text{sw} \rightarrow \hat{s}}) = \varsigma \left(V(\hat{s}) - \frac{N_{\text{fb}}}{2} \right) + \frac{N_{\text{fb}}}{2}. \quad (5.35)$$

- The focus of action $a_{\text{dt} \rightarrow \hat{s}}$ (performing a detection in frequency band \hat{s}) is exploration. Therefore, it is preferable to perform detections in frequency bands, where the reliability of the state-values are low. This can be achieved by the following mapping of state-values to preferences:

$$p(s, a_{\text{dt} \rightarrow \hat{s}}) = (1 - \varsigma) \cdot V(s). \quad (5.36)$$

The modified actor–critic structure for the extended model is visualized in Figure 5.15 and the corresponding learning algorithm is given in procedural form in Algorithm 5.8.

For the simulations the same parameters were used as for the basic model, making the simulation results comparable. Figure 5.14 shows that the overall performance is better than for the basic model without out-of-band detections.

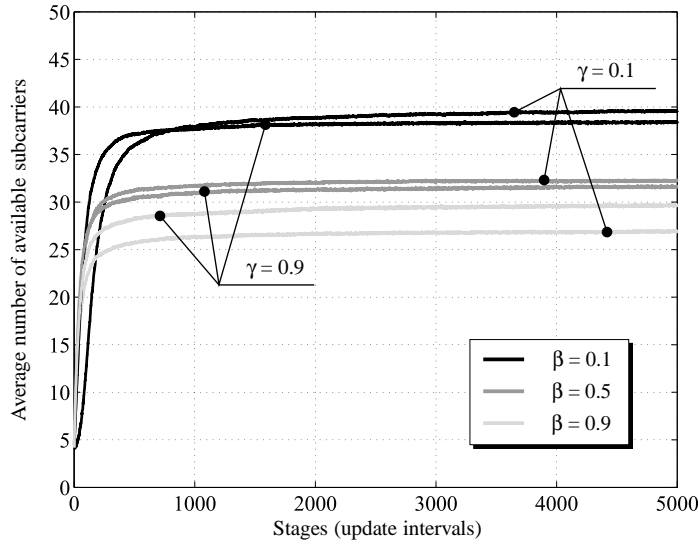


Figure 5.14 Average number of received rewards (available subcarriers) depending on the stage for the extended model using an enhanced actor–critic method.

Algorithm 5.8: Learning algorithm for the extended model using a modified temporal difference approach.

```

1 foreach episode do
2   Initialize s
3   repeat for each step of episode:
4     Choose a from s using policy derived from V
5     Take action a, observe r, s'
6     Update reliability vector:  $\varsigma_t \leftarrow \varsigma_t + \alpha_1 [d - \varsigma_t]$ 
7     Update estimation of resources:
8        $\hat{u}_n(t+1) = \hat{u}_n(t) + \alpha_2 [u_n(t+1) - \hat{u}_n(t)]$ 
9     Calculate weighted estimations:  $\Upsilon_n(t) = \varsigma_n(t) (\hat{u}_n(t) - \frac{N_{fb}}{2}) + \frac{N_{fb}}{2}$ 
10    Calculate TD-error:  $\delta_t = r_{t+1} + \gamma \Upsilon(s_{t+1}) - V(s_t)$ 
11    Update state-values:  $V(s_t) \leftarrow V(s_t) + \beta \delta_t$ 
12    Update preferences
13    Update policy
14  until s is terminal
15 end

```

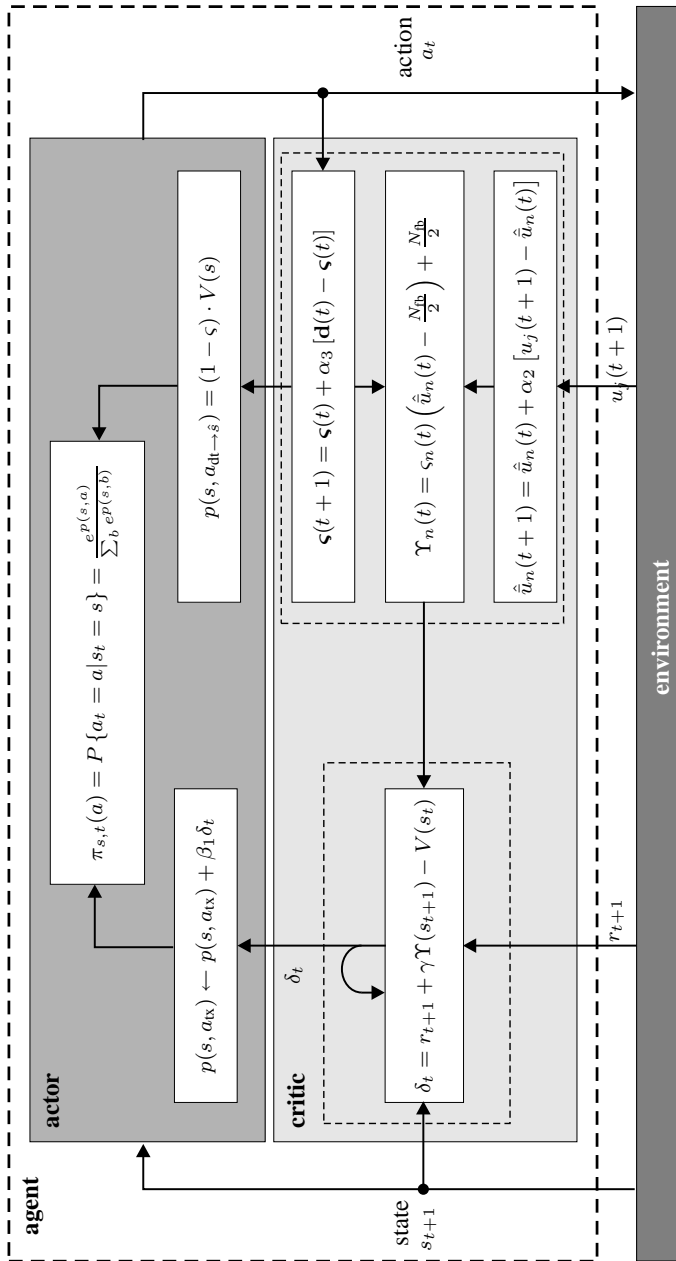


Figure 5.15 Modified actor-critic structure for extended model.

6 Conclusion

Overlay systems based on OFDM are a good candidate for increasing the efficiency in spectrum use in frequency bands where a PU system is already deployed and an opportunistic spectrum access is not possible. Since the SU system is responsible for ensuring a smooth coexistence of both systems in the same frequency band, it has to provide all necessary features for reducing mutual interference and for finding transmission opportunities at the same time. In this work, the properties of PU systems were discussed and guidelines for designing overlay systems were derived. Several aspects of OFDM based overlay systems were investigated, including an optimization of the SU system's parameter configuration and a discussion of the available resources for an SU system covering two different spectrum occupation zones.

One major contribution of this work is a novel method for a distributed detection for SU systems operating in ad hoc mode. Based on a model from graph theory and a model for energy detection, the average network detection probability and false alarm probability were defined. Then theoretical bounds for the minimum necessary node density were derived for still achieving a required network detection probability. In a more realistic scenario, additionally the signaling of the detection results was considered. Due to the lacking central access point, signaling has to be performed in a distributed manner. This can be efficiently achieved by the proposed architecture where each node first performs a single detection phase, which is then followed by a pair of boosting and collection phases. The order is chosen randomly so that each node contributes to the distributed detection evaluation of all of its neighbors but also receives enough information about their detection results. Furthermore, the relation of the single detection phase and the signaling phase regarding their duration and its impact on the overall performance was investigated. The distributed detection approach requires that all SUs are silent during the detection period so that the detection results are not interfered by signals from the SU system. Especially in an ad hoc scenario this requires a coordination of the detection phases, which is provided by the proposed AHOMAC protocol. Using a slot-based frame structure with a dedicated detection phase and synchronization phase it allows the detection of a PU system (overlay property) as well as the detection of other nodes in the neighborhood (ad hoc property).

Furthermore, a multiband scenario was considered, consisting of several frequency bands with a different average spectrum occupation. The SU system can only op-

erate in a single frequency band in which it has to perform accurate measurements for detecting the PU system's spectrum access. Nevertheless, it has to observe the average spectrum occupation in the other frequency bands to be able to find a better opportunity for transmissions. Therefore, it has to find a trade-off between exploring the other frequency bands, thereby gaining new information, and exploiting the already available information. In this work it has been shown that this problem can be modeled as a learning problem and solved by reinforcement learning methods. In particular, depending on the assumptions regarding the signaling overhead, three different problem formulations were proposed. The formulation similar to the n -armed bandit problem serves as a reference model, since it assumes no signaling overhead. The basic model, however, includes signaling overhead and is formulated as an MDP. The proposed solution strategy applies an actor-critic approach, since this allows an on-line implementation without knowing the complete system dynamics. Simulations show that the maximum performance is approximated more slowly with an increased signaling overhead. Finally, the basic model was extended by a further action allowing out-of-band detection phases and a reliability function. With this additional action the overall performance can be increased. Nevertheless, one drawback of the discussed reinforcement learning approaches is the number of parameters that have to be adjusted for achieving the optimum performance, including for example several step-size parameters and the discount factor. However, in future work this can be handled for example by a meta-learning approach where on a second level these parameters are adjusted by a learning algorithm after setting them to good start values.

Acronyms

8-PSK	octal phase shift keying	17
ACK	acknowledgement	102
ATC	air traffic control	14
AWGN	additive white Gaussian noise	24
AHOMAC	ad hoc overlay MAC	74
CBR	constant bit rate	74
CDMA	code division multiple access	12
ComSoc	Communications Society	3
CSMA/CA	carrier sense multiple access with collision avoidance	13
CTR	critical transmission range	44
CTS	clear to send	78
DySPAN	International Symposium on Dynamic Spectrum Access Networks .	3
DARPA	Defense Advanced Research Projects Agency	41
DME	distance measuring equipment	
DCF	distributed coordination function	74
DSA	dynamic spectrum access	2
DSB-AM	double sideband amplitude modulation	14
DSL	digital subscriber line	17
FCC	Federal Communications Commission	1
FDD	frequency division duplex	16
FDMA	frequency division multiple access	13
FFT	fast Fourier transform	13
GMSK	Gaussian minimum-shift keying	17
GPRS	general packet radio service	2
GSM	Global System for Mobile communication	2
IEEE	Institute of Electrical and Electronics Engineers	3
ISM	industrial, scientific and medical	8
LR	likelihood ratio	25

MAC	medium access control	5
MAN	metropolitan area network	17
MANET	mobile ad hoc network	41
MC-CDMA	multi-carrier code division multiple access	13
MDP	Markov decision process	4
MDR	multiservices dynamic reservation	74
MIMO	multiple input multiple output	6
NLOS	no line of sight	24
OFDM	orthogonal frequency division multiplex	3
PAN	private area network	41
PHY	physical	5
PRMA	packet reservation multiple access	74
PU	primary user	2
pdf	probability density function	25
QAM	quadrature amplitude modulation	18
QPSK	quadrature phase shift keying	18
ROC	receiver operating characteristic	47
RTS	request to send	76
SARSA	state-action-reward-state-action	98
SNR	signal-to-noise ratio	31
SOV	spectrum occupancy vector	13
SSR	secondary surveillance radar	16
SU	secondary user	2
SU-AP	secondary user access point	24
TACAN	tactical air navigation	16
TAKOKO	Techniken, Algorithmen und Konzepte für COFDM Systeme zur Koexistenz mit autorisierten Systemen im selben Frequenzband	14
TD	temporal-difference	95
TDD	time division duplex	18
TDMA	time division multiple access	12

TETRA	terrestrial trunked radio	43
TV	television	6
UWB	ultra-wideband	2
VANET	vehicular ad hoc network	42
VBR	variable bit rate	74
VHF	very high frequency	14
WLAN	wireless local area network	15

Bibliography

- [1] IEEE 802.22 working group. www.ieee802.org/22.
- [2] IEEE standards coordinating committee 41. www.scc41.org.
- [3] ETSI terrestrial trunked radio (TETRA); voice plus data (V+D); part 1: General network design. *ETSI Std EN 300 392-1*, 2004.
- [4] IEEE standard for local and metropolitan area networks part 16: Air interface for fixed broadband wireless access systems. *IEEE Std 802.16-2004*, 2004.
- [5] Bluetooth specification version 2.1. www.bluetooth.com, 2007.
- [6] IEEE standard for local and metropolitan area networks part 11: Wireless LAN MAC and PHY specifications. *IEEE Std 802.11-2007*, 2007.
- [7] Norman Abramson. The throughput of packet broadcasting channels. *IEEE Trans. Commun.*, 25(1):117–128, 1977.
- [8] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. A survey on sensor networks. *IEEE Commun. Mag.*, 40(8):102–114, Aug. 2002.
- [9] Ian F. Akyildiz, Xudong Wang, and Weilin Wang. Wireless mesh networks: a survey. *Computer Networks*, 47(4):445–487, March 2005.
- [10] B-VHF project. Interference on the B-VHF Overlay System. Technical Report D-09, www.b-vhf.org, October 2005.
- [11] André N. Barreto and Simeon Furrer. Adaptive bit loading for wireless OFDM systems. In *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, volume 2, pages G–88–G–92, 2001.
- [12] Lars Berlemann, Stefan Mangold, and Bernhard H. Walke. Policy-based reasoning for spectrum sharing in radio networks. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 1–10, 2005.
- [13] Ulrich Berthold, Sinja Brandes, Friedrich K. Jondral, and Michael Schnell. Performance of OFDM-CDMA overlay systems considering inaccurate allocation vectors. In *Proc. of the 10th International OFDM Workshop*, pages 230–234, Hamburg, Germany, August 2005.
- [14] Ulrich Berthold, Sinja Brandes, Friedrich K. Jondral, and Michael Schnell. A framework for crosslayer optimization in OFDM based overlay systems. In *Proceedings of the 11th International OFDM-Workshop*, pages 253–257, Hamburg, Germany, August 2006.

- [15] Ulrich Berthold, Sinja Brandes, Friedrich K. Jondral, and Michael Schnell. OFDM based overlay systems - a promising approach for enhancing spectral efficiency. *IEEE Commun. Mag.*, December 2007.
- [16] Ulrich Berthold, Sinja Brandes, Michael Schnell, and Friedrich K. Jondral. On focus: OFDM based overlay scenarios. In *Proceedings of the 2006 IST Mobile and Wireless Communications Summit*, Myconos, Greece, June 2006. CDROM.
- [17] Ulrich Berthold, Fangwen Fu, Mihaela van der Schaar, and Friedrich K. Jondral. Detection of spectral resources in cognitive radios using reinforcement learning. In *IEEE 2nd International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, October 2008.
- [18] Ulrich Berthold, Holger Heimpel, and Friedrich K. Jondral. Coordination of allocation measurements in OFDM based ad hoc overlay systems. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, volume 4, pages IV—1365–IV—1368, 2007.
- [19] Ulrich Berthold and Friedrich K. Jondral. Generalized system model of an overlay environment. In *Proceedings of the 2005 Software Defined Radio Forum Technical Conference (SDR '05), Orange County (CA), USA*. CDROM, November 2005.
- [20] Ulrich Berthold and Friedrich K. Jondral. Guidelines for designing OFDM overlay systems. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 626–629, November 2005.
- [21] Ulrich Berthold and Friedrich K. Jondral. Distributed detection in OFDM based ad hoc overlay systems. In *IEEE 67th Vehicular Technology Conference, VTC Spring*, pages 1666–1670, 2008.
- [22] Dimitri P. Bertsekas and Robert G. Gallager. *Data networks*. Prentice-Hall International, 2. ed. edition, 1992.
- [23] Christian Bettstetter. On the Connectivity of Ad Hoc Networks. *The Computer Journal*, 47(4):432–447, 2004.
- [24] Béla Bollobás. *Modern graph theory*. Springer, 1998.
- [25] Sinja Brandes and Ulrich Berthold. TAKOKO – Technischer Bericht 1: Definition der Szenarien, March 2005.
- [26] Sinja Brandes, Ulrich Berthold, Michael Schnell, and Friedrich K. Jondral. OFDM based overlay systems – design challenges and solutions. In *18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2007.
- [27] Sinja Brandes, Ivan Cosovic, and Michael Schnell. Reduction of out-of-band radiation in OFDM based overlay systems. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 662–665, 2005.
- [28] Sinja Brandes, Ivan Cosovic, and Michael Schnell. Reduction of out-of-band radiation in OFDM systems by insertion of cancellation carriers. *IEEE Commun. Lett.*, 10(6):420–422, 2006.

- [29] Sinja Brandes and Michael Schnell. Mitigation of dynamically changing NBI in OFDM based overlay systems. In *International Multi-Carrier Spread Spectrum Workshop (MC-SS 2007)*, pages 117–126, Herrsching, Germany, May 2007.
- [30] Raffaele Bruno, Marco Conti, and Enrico Gregori. Mesh networks: Commodity multihop ad hoc networks. *IEEE Commun. Mag.*, 43(3):123–131, March 2005.
- [31] Dennis Burgkhardt, Ivan Cosovic, and Friedrich K. Jondral. Dynamic spectrum allocation by hierarchical resource trading. In *IEEE 67th Vehicular Technology Conference, VTC Spring*, May 2008.
- [32] Christos G. Cassandras and Stéphane Lafortune. *Introduction to discrete event systems*. Springer, 2. ed. edition, 2008.
- [33] John M. Chapin and William H. Lehr. Time-limited leases for innovative radios. In *IEEE 2nd International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 606–619, 2007.
- [34] John M. Chapin and William H. Lehr. Time-limited leases in radio systems. *IEEE Commun. Mag.*, 45(6):76–82, 2007.
- [35] Fan R. Chung. *Spectral graph theory*. American Mathematical Society, repr. with corr., 2. [pr.] edition, 1997.
- [36] Carlos Cordeiro, Kiran Challapali, Dagnachew Birru, and Sai Shankar. IEEE 802.22: the first worldwide wireless standard based on cognitive radios. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 328–337, 2005.
- [37] David Culler, Deborah Estrin, and Mani Srivastava. Overview of sensor networks. *Computer*, 37(8):41–49, Aug. 2004.
- [38] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes*. Springer, 1988.
- [39] Grit Denker, Daniel Elenius, Rukman Senanayake, Mark-Oliver Stehr, and David Wilkins. A policy engine for spectrum sharing. In *IEEE 2nd International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 55–65, 2007.
- [40] Stuart Dreyfus. Richard Bellman on the birth of dynamic programming. *Operations Research*, 50(1):48–51, January 2002.
- [41] ETSI. TS 100 573, Digital cellular telecommunications system (Phase 2+); Physical layer on the radio path; General description, 1999.
- [42] Federal Aviation Administration. United States National Aviation Standard for the Very High Frequency Omnidirectional Radio Range (VOR)/ Distance Measuring Equipment (DME)/ Tactical Air Navigation (TACAN) Systems, 1982. AC No. 00-31A.
- [43] Fangwen Fu and Mihaela van der Schaar. A new theoretic framework for cross-layer optimization. In *Proc. IEEE Int. Conf. on Image Process. 2008 (ICIP 2008)*, to appear 2008.

- [44] Izrail' S. Gradshteyn and Iosif M. Ryzhik. *Table of integrals, series, and products*. Academic Press, San Diego, 6 edition, 2000.
- [45] Lajos Hanzo. *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. Wiley, 2003.
- [46] Lajos Hanzo, Matthias Münster, Byoung-Jo Choi, and Thomas Keller. *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*. Wiley [u.a.], 2003.
- [47] Hannes Hartenstein and Kenneth P. Laberteaux. Topics in ad hoc and sensor networks – a tutorial survey on vehicular ad hoc networks. *IEEE Commun. Mag.*, 46(6):164–171, June 2008.
- [48] Dale N. Hatfield and Philip J. Weiser. Property rights in spectrum: Taking the next step. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 43–55, 2005.
- [49] Jörg Hillenbrand, Timo Weiß, and Friedrich K. Jondral. Calculation of detection and false alarm probabilities in spectrum pooling systems. *IEEE Commun. Lett.*, 9(4):349–351, 2005.
- [50] Wendong Hu, Daniel Willkomm, Murad Abusubaih, James Gross, George Vlantis, Mario Gerla, and Adam Wolisz. Dynamic frequency hopping communities for efficient IEEE 802.22 operation. *IEEE Communications Magazine*, 45(5):80–87, May 2007.
- [51] Friedrich K. Jondral. Cognitive radio: A communications engineering view. *IEEE Wireless Commun. Mag.*, 14(4):28–33, August 2007.
- [52] Robert E. Kahn, Steven A. Gronemeyer, Jerry Burchfiel, and Ronald C. Kunzelman. Advances in packet radio technology. *Proc. IEEE*, 66(11):1468–1496, Nov. 1978.
- [53] Clemens Klöck. *Auction-based Medium Access Control*. PhD thesis, Universität Karlsruhe (TH), Institut für Nachrichtentechnik, 2007.
- [54] Mieczyslaw Kokar, Donald Hillman, Shujun Li, Bruce Fette, Preston Marshall, Mark Cummings, Todd Martin, and John Strassner. Towards a unified policy language for future communication networks: A process. In *IEEE 3rd International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, 2008.
- [55] Mark McHenry. NSF spectrum occupancy measurements. Technical report, The Shared Spectrum Company, www.sharedspectrum.com/?section=nsf_measurements, 2005.
- [56] Mark A. McHenry, Peter A. Tenhula, Dan McCloskey, Dennis Roberson, and Cynthia Wood. Chicago spectrum occupancy measurements & analysis and a long-term proposal. In *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*. www.wtapas.org, August 2006.
- [57] Shridhar M. Mishra, Anant Sahai, and Robert W. Brodersen. Cooperative sensing among cognitive radios. In *IEEE International Conference on Communications, ICC*, volume 4, pages 1658–1663, June 2006.

- [58] Richard van Nee and Ramjee Prasad. *OFDM for wireless multimedia communications*. Artech House, 2000.
- [59] Marc P. Olivieri, Greg Barnett, Alex Lackpour, Albert Davis, and Phuong Ngo. A scalable dynamic spectrum allocation system with interference mitigation for teams of spectrally agile software defined radios. In *IEEE 1st International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, pages 170–179, 2005.
- [60] Mengüç Öner and Friedrich K. Jondral. Cyclostationarity-based methods for the extraction of the channel allocation information in a spectrum pooling system. In *Proc. IEEE Radio and Wireless Conference*, pages 279–282, 2004.
- [61] Mengüç Öner and Friedrich K. Jondral. On the extraction of the channel allocation information in spectrum pooling systems. *IEEE J. Select. Areas Commun.*, 25(3):558–565, 2007.
- [62] Ian Oppermann, Matti Hämäläinen, and Jari Iinatti, editors. *UWB: Theory and Applications*. John Wiley and Sohns, 2004.
- [63] Panagiotis Papadimitratos, Srivatsan Sankaranarayanan, and Amitabh Mishra. A bandwidth sharing approach to improve licensed spectrum utilization. *IEEE Commun. Mag.*, 43(12):S10–S14, December 2005.
- [64] Jon M. Peha. Approaches to spectrum sharing. *IEEE Commun. Mag.*, 43:10–12, February 2005.
- [65] Mathew D. Penrose. *Random geometric graphs*. Oxford University Press, 2004.
- [66] Martin L. Puterman. *Markov decision processes*. Wiley Interscience, 2005.
- [67] Theodore S. Rappaport. *Wireless communications*. Prentice Hall, 2. edition, 2002.
- [68] D. Raychaudhuri and N. D. Wilson. ATM-based transport architecture for multiservices wireless personal communication networks. *IEEE Journal on Selected Areas in Communications*, 12:1401 – 1414, October 1994.
- [69] Tobias Renk, Clemens Klöck, and Friedrich K. Jondral. A cognitive approach to the detection of spectrum holes in wireless networks. In *4th IEEE Consumer Communications and Networking Conference CCNC 2007*, pages 1118–1122, 2007.
- [70] Jeremy Rifkin. *The European dream : how Europe's vision of the future is quietly eclipsing the American dream*. Tarcher, August 2004.
- [71] Gavin A. Rummery and Mahesan Niranjan. On-line q-learning using connectionist systems. Technical report, 1994.
- [72] Konstantin A. Semendjaev. *Taschenbuch der Mathematik*. Deutsch, 7., vollst. überarb. u. erg. edition, 2008.
- [73] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning*. MIT Press, [repr.] edition, [circa 2004].

- [74] Gerald Tesauro. TD-gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.*, 6(2):215–219, March 1994.
- [75] Gerald Tesauro. Temporal difference learning and TD-gammon. *Commun. ACM*, 38(3):58–68, 1995.
- [76] Zhi Tian and Georgios B. Giannakis. Compressed sensing for wideband cognitive radios. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, volume 4, pages IV–1357–IV–1360, 2007.
- [77] George V. Tsoulos, editor. *MIMO system technology for wireless communications*. CRC Press, 2006.
- [78] Harry Urkowitz. *Signal theory and random processes*. Artech House, Dedham, 1983.
- [79] Christopher J. C. H. Watkins. *Learning from Delayed Rewards*. PhD thesis, King’s College, Cambridge, UK, 1989.
- [80] Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.
- [81] Timo Weiß. *OFDM-basiertes Spectrum Pooling*. PhD thesis, Universität Karlsruhe (TH), 2004.
- [82] Timo Weiß, Jörg Hillenbrand, Albert Krohn, and Friedrich K. Jondral. Efficient signaling of spectral resources in spectrum pooling systems. In *10th Symposium on Communications and Vehicular Technology, SCVT*, 2003.
- [83] Timo Weiß, Jörg Hillenbrand, and Friedrich K. Jondral. A diversity approach for the detection of idle spectral resources in spectrum pooling systems. In *Proc. of the 48th International Scientific Colloquium, ISC 2003, Illmenau, Germany*, 2003.
- [84] Timo Weiß and Friedrich K. Jondral. Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency. *IEEE Commun. Mag.*, 42:8–14, 2004.
- [85] Lin Xu, Ralf Tonjes, Toni Paila, Wolfgang Hansmann, Matthias Frank, and Markus Albrecht. DRiVE-ing to the internet: Dynamic radio for IP services in vehicular environments. In *25th Annual IEEE Conference on Local Computer Networks, LCN 2000*, pages 281–289, 2000.
- [86] Takefumi Yamada, Dennis Burgkhardt, Ivan Cosovic, and Friedrich K. Jondral. Resource distribution approaches in spectrum sharing systems. *EURASIP Journal on Wireless Communications and Networking*, 2008:1–15, 2008.
- [87] Qing Zhao and Brian M. Sadler. A survey of dynamic spectrum access. *IEEE Signal Processing Mag.*, 24(3):79–89, May 2007.
- [88] Qing Zhao, Lang Tong, Ananthram Swami, and Yunxia Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE J. Select. Areas Commun.*, 25(3):589–600, April 2007.

Theses

- Aleksey Burlak,
Yannick Dieudonné,
Min Wen: *Aufbau und Inbetriebnahme eines Low Cost Video
Konferenzsystems*
Teamproject, February 3, 2005
- Holger Heimpel: *Verfahren zur Koordination der Frequenzlückende-
tektion in einem Ad-Hoc Overlay-System*
Diplomarbeit, December 9, 2005
- Masoud Hamid: *Untersuchung von Multi-Carrier-CDMA in Over-
lay Systemen*
Master thesis, December 20, 2005
- Ankit Kaushik: *Performance Analysis of a MAC Protocol for
OFDM based Overlay Systems in Ad Hoc Mode*
Master thesis, July 14, 2007
- Noha El Gemayel: *Untersuchung und Bewertung von IEEE 802.11s
und WiMAX im Mesh-Mode als Overlay-Systeme*
Studienarbeit, December 15, 2008

Index

- χ^2 -distribution, 26
- policy description language, 9
- ALOHA, 6
- blocked spectrum, 6
- boosting protocol, 48
- boosting signal, 24
- compressed sensing, 10
- connected graph, 44
- coverage area, 36, 43
- degree of a node, 44
- detection probability, 25
 - single user, 26
- detector model, 24
- edge, 43
- efficiency
 - communication network, 6
 - multiple coexisting communication networks, 6
 - point-to-point transmission, 5
 - spectral efficiency, 6
- Erlang B formula, 28
- false alarm probability, 25
 - single user, 26
- graph theory, 43
- IEEE
 - 802.16, 18
 - 802.22, 11
- 1900, 11
- isolated node, 44
- likelihood ratio test, 25
- machine learning, 82
- Markov chain, 27
- mean power, 25
- neighborhood of a node, 44
- Neyman-Pearson, 25
- node, 43
- node detection, 75
- Poisson distribution, 45
- policy, 83
- random geometric graphs, 44
- reinforcement learning, 82
- reward function, 83
- secondary spectrum market, 9
- service rate, 27
- softmax action selection, 100
- spectrum auctions, 10
- spectrum commons, 8
- spectrum management, 10
- spectrum occupation vector, 13
- spectrum occupation zone, 36
- spectrum sensing, 10
- system area, 44
- task
 - continuing task, 86
 - episodic task, 86
- time/frequency plane, 22

traffic model

 simple model, 27

transmission range, 43

value function, 83

Förderung

Die dieser Dissertation zu Grunde liegenden Arbeiten wurden teilweise im Rahmen der Projekte TAKOKO (Techniken, Algorithmen und Konzepte für COFDM Systeme zur Koexistenz mit autorisierten Systemen im selben Frequenzband, Förderkennzeichen: JO258/10-2) und OOS (OFDM-basierte Overlay-Systeme zur dynamischen und effizienten Nutzung des Spektrums, Förderkennzeichen: JO258/13-1) durch die Deutsche Forschungsgemeinschaft (DFG) gefördert. Das Projekt TAKOKO war in das Schwerpunktprogramm TakeOFDM (Techniken, Algorithmen und Konzepte für zukünftige COFDM Systeme) eingebettet.

Der im Rahmen dieser Dissertation erfolgte Forschungsaufenthalt am Multimedia Communications and Systems Laboratory der University of California, Los Angeles (UCLA), CA, USA, wurde teilweise durch das Karlsruhe House of Young Scientists (KHYS) des Karlsruhe Institute of Technology (KIT) gefördert.

Curriculum Vitae

Personal Data

Name	Ulrich Berthold
Date of birth	January 18, 1976
Place of birth	Mainz
Nationality	German

School Education

1982-1984	Primary School in Plainfield, NJ, USA
1984-1986	Primary School in Mainz, Germany
1986-1995	Theresianum Mainz, Gymnasium des Johannesbundes

Academic Education

1996-2003	Universität Karlsruhe (TH), Studies in electrical engineering with focus on communication engineering
2004-2009	PhD studies at Universität Karlsruhe (TH), Institut für Nachrichtentechnik

Work History

01/2004-04/2004	Software engineer at fun communications GmbH, Karlsruhe, Germany
since 05/2004	Research assistant (wissenschaftlicher Mitarbeiter) at Universität Karlsruhe (TH), Institut für Nachrichtentechnik