KTH Electrical Engineering

Distributed Detection in Cognitive Radio Networks

AHTI AINOMÄE

Licentiate Thesis in Electrical Engineering
Tallinn, Estonia; Stockholm, Sweden 2017

KTH, School of Electrical Engineering
Department of Signal Processing
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2017:109
ISSN 1653-5146
ISBN 978-91-7729-515-0

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av licentiatexamen i Elektroteknik fredagen den 28 September 2017 klockan 13:00 i hörsal V3, Teknikringen 72, Stockholm.

Tryck: Universitetsservice US AB

To my parents

Abstract

One of the problems with the modern radio communication is the lack of available radio frequencies. Recent studies have shown that, while the available licensed radio spectrum becomes more occupied, the assigned spectrum is significantly underutilized. To alleviate the situation, cognitive radio (CR) technology has been proposed to provide an opportunistic access to the licensed spectrum areas. Secondary CR systems need to cyclically detect the presence of a primary user by continuously sensing the spectrum area of interest. Radiowave propagation effects like fading and shadowing often complicate sensing of spectrum holes. When spectrum sensing is performed in a cooperative manner, then the resulting sensing performance can be improved and stabilized.

In this thesis, two fully distributed and adaptive cooperative Primary User (PU) detection solutions for CR networks are studied.

In the first part of this thesis we study a distributed energy detection scheme without using any fusion center. Due to reduced communication such a topology is more energy efficient. We propose the usage of distributed, diffusion least mean square (LMS) type of power estimation algorithms with different network topologies. We analyze the resulting energy detection performance by using a common framework and verify the theoretical findings through simulations.

In the second part of this thesis we propose a fully distributed detection scheme, based on the largest eigenvalue of adaptively estimated correlation matrices, assuming that the primary user signal is temporally correlated. Different forms of diffusion LMS algorithms are used for estimating and averaging the correlation matrices over the CR network. The resulting detection performance is analyzed using a common framework. In order to obtain analytic results on the detection performance, the adaptive correlation matrix estimates are approximated by a Wishart distribution. The theoretical findings are verified through simulations.

Keywords: Cognitive Radio, distributed estimation, distributed detection, Diffusion LMS, Diffusion Networks, Adaptive Networks, Spectrum Sensing, Energy Detection, Random Matrix, Largest Eigenvalue Detection.

Sammanfattning

Ett av de framtida problemen med modern radiokommunikation är bristen på tillgängliga radiofrekvenser. Tidigare studier har visat att medan det tillgängliga licensierade radiospektrumet blir mer och mer upptaget, är det tilldelade spektret betydligt underutnyttjat. För att lindra situationen har kognitiv radio (CR) föreslagits för att ge en opportunistisk tillgång till de licensierade spektrumområdena. Sekundära CR-system måste regelbundet detektera närvaron av en primär användare genom att kontinuerligt känna av det intressanta spektrumområdet. Sökandet efter spektrumhål kompliceras ofta av radioutbredningsfenomen såsom fädning och skuggning. När spektrumavkänningen utförs i samarbete mellan flera noder, kan den resulterande detektionsprestandan förbättras och stabiliseras.

I denna avhandling studeras två fullt distribuerade och adaptiva kooperativa lösningar för att detektera primära användare (PU) i CR-nätverk.

I den första delen av avhandlingen studerar vi ett distribuerat energidetektionssystem utan användning av något fusionscenter. På grund av minskad kommunikation är en sådan topologi mer energieffektiv. Vi föreslår användningen av energidetektion baserad på distribuerad, diffusions-LMS med olika nätverkstopologier och studerar resulterande prestanda. Vi analyserar den resulterande energidetekteringsprestandan genom att använda ett gemensamt ramverk och verifierar de teoretiska resultaten genom simuleringar.

I den andra delen av avhandlingen föreslår vi ett helt distribuerat detekteringsschema baserat på det största egenvärdet för adaptivt skattade korrelationsmatriser, under förutsättningen att den primära användarsignalen är temporärt korrelerad. Olika former av diffusions-LMS-algoritmer används för att uppskatta och medelvärdesbilda korrelationsmatriserna över CR-nätverket. Den resulterande detektionsprestandan analyseras med hjälp av ett gemensamt ramverk. För att erhålla analytiska resultat på detektionsprestandan approximeras de adaptiva korrelationsuppskattningarna av en Wishart-fördelning. De teoretiska resultaten verifieras genom simuleringar.

Nyckelord: Kognitiv radio, distribuerad uppskattning, distribuerad Detektion, diffusion LMS, diffusionsnät, adaptiva nätverk, spektrum Sensing, Energy Detection, Slumpmässig Matrix, Största Eigenvalue Detection

Acknowledgments

First and foremost, I am grateful to my academical advisors Prof. Mats Bengtsson and Prof. Tõnu Trump for providing me with an opportunity to study jointly in the Department of Signal Processing at KTH Royal Institute of Technology and in the Department of Radio and Communication Engineering at Tallinn University of Technology (TUT). I am thankful for all the freedom they gave me to explore the research topics and directions. Despite their busy schedule, they were always available for discussion. I am grateful for Prof. Tõnu Trump for leading me in the research area of statistical and adaptive signal processing and for establishing the cooperation between TUT and KTH for my joint studies. I have been always impressed with the research ideas of Prof. Tõnu Trump and the feedback for my papers he provided. I highly evaluate the research ideas, recommendations for driving the ongoing research and all the support from Prof. Mats Bengtsson which helped me a lot to achieve the current work results. Additionally I would like to thank Prof. Peter Händel for supporting me with the studies in the Department of Signal Processing at KTH. Also I would like to thank the Thomas Johann Seebeck Department of Electronic at TUT for supporting me in the preparation period of my Licentiate Thesis.

I have been fortunate to collaborate with brilliant colleagues like Älla Tarighati, Vijaya Yajnanarayana, Klas Magnusson, Rasmus Brandt, Martin Sundin, Senay Amanuel Negusse, Marie Maross, Pol Del Aguila Pla, Arun Venkitaraman, Nima Najari Moghadam, Arash Owrang, Ehsan Olfat during my studies in KTH and Sander Ulp, Maksim Butsenko, Julia Berdnikova during my studies at TUT. Many thanks for collaborating with me. I also want to thank Prof. Peter Händel, Prof. Magnus Jansson, Associate Prof. Tobias Oechtering, Associate Prof. Joakim Jaldén, Dr. Satyam Dwivedi for teaching me courses in statistical signal processing. These topics helped me a lot in my research.

I would like to thank Prof. Mats Bengtsson for the Swedish translation of the abstract, Prof. Mats Bengtsson and Prof. Tõnu Trump for the proof-reading of the thesis and Associate Prof. Tobias Oechtering for performing the quality check of the thesis. Many thanks to Tove Schwartz and Raine Tiivel for assistance in administrative affairs. I am grateful to Karolin Trump and Samuel Njie for all the help regarding the accommodation and supply issues during my stay in Stockholm.

I wish to thank Associate Prof. Olav Tirkkonen of Aalto University in Finland for taking time to act as opponent to this thesis.

Finally I would like to express my deepest gratitude to my family for their love and support.

Ahti Ainomäe
Tallinn, September 2017

Contents

Nomenclature

Abbreviations and Acronyms

3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
ATC	Adapt and The Combine
AWGN	Additive White Gaussian Noise
CDF	Cumulative Distribution Function
CM	Correlation Matrix
CR	Cognitive Radio
CSCG	Circular Symmetric Complex Gaussian Distribution
CTA	Combine and Then Adapt
dB	decibel
ED	Energy Detection
FC	Fusion Center
GSM	Global System for Mobile Communications
IEEE	Institute of Electrical and Electronics Engineers
IID	Independent and Identically Distributed
IoT	Internet of Things
LE	Largest Eigenvalue Detection
LMS	Least Mean Square
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Squared Error
MSE	Mean of the Squared Error
OFDM	Orthogonal Frequency Division Multiplexing
PD	Probability of Detection

PDF	Probability Density Function
PFA	Probability of False Alarm
PSD	Power Spectral density
PU	Primary User in a Cognitive Radio Network
RF	Radio Frequency
RLS	Recursive Least Squares
SNR	Signal to Noise Ratio
WSN	Wireless Sensor Networks

Notations

a	Scalars are written in normal font
$\|a\|$	The absolute value of the scalar a
$\mathbf{a}$	Vectors are written in lower-case bold font
$\\|\mathbf{x}\\|_p$	p-norm of $\mathbf{x}$
$\mathbb{R}^N$	N-dimensional real field
$\mathbb{C}^N$	N-dimensional complex field
$\\|\mathbf{x}\\|$	2-norm of $\mathbf{x}$
$\mathbf{A}$	Matrices are written in upper-case bold font
$\mathbf{a}_{i_k,n}$	The nth column of the matrix $\mathbf{A}_{i_k}$
$\mathbf{A}^{-1}$	Matrix Inverse of $\mathbf{A}$
$[\mathbf{A}]_{i,j}$	Element corresponding to i row and j column
$\mathbf{A}^{\mathrm{T}}$	Matrix transpose
$\mathbf{A}^{*}$	Conjugate transpose
$\mathbf{A}^{\mathbf{H}}$	Hermitian transpose
$\otimes$	Kronecker product
$\mathbf{I}_n$	The identity matrix of size $n \times n$
$\lambda_{\max}(\mathbf{A})$	The eigenvalue of the matrix $\mathbf{A}$ with the largest magnitude
$\lambda_{\min}(\mathbf{A})$	The eigenvalue of the matrix $\mathbf{A}$ with the smallest magnitude
$\Pr(\mathcal{E})$	Probability of the event $\mathcal{E}$
$p(x)$	PDF of x
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	The circularly-symmetric complex Gaussian probability distribution with mean $\mathbf{m}$ and covariance $\mathbf{C}$
$\mathcal{CW}_M(N, \boldsymbol{\Sigma})$	Complex Wishart distribution of dimension M with degree of freedom N and population covariance matrix $\boldsymbol{\Sigma}$
$a \sim b$	The random variable a is distributed according to the probability distribution b

$\mathrm{E}[\cdot]$	Expectation of a random variable
$\mathrm{Var}[\cdot]$	Variance of a random variable
$\mathrm{Cov}[\cdot]$	Covariance of a random vector
$\mathrm{vec}[\cdot]$	Vectorization of a matrix
$\det[\cdot]$	Determinant of a matrix
$\mathrm{diag}[\cdot]$	Vector of diagonal elements of a matrix
$\mathrm{Tr}[\cdot]$	Trace of a matrix
$\{a, b, c\}$	Set with elements a, b and c
$[a, b]$	Closed interval between a and b
$l \in \mathcal{N}_k / \{k\}$	Variable l belongs to the set of $\mathcal{N}_k$ by excluding the element of k from the set
$\mathbf{Q}(\cdot)$	Tail Probability of standard normal function
argmin	The argument that achieves minimum
$\binom{n}{k}$	The (n, k)th binomial coefficient
$\mathcal{C}$	The set of complex numbers
$\mathrm{Re}(c)$	Real part of the complex number $c \in \mathcal{C}$
$\mathrm{Im}(c)$	Imaginary part of the complex number $c \in \mathcal{C}$

List of Figures

Chapter 1

Introduction

Future communication networks will have seamless and ubiquitous connectivity among several communicating devices using different radio technologies. In the year 2021, it is predicted that there will be 16 billion devices that will be connected [ER116]. These devices could include cell phones, TVs, computers, tablets, etc. Wireless sensor networks play an important role in the future of Internet of Things systems. Several applications as Smart Grids, Smart Homes, Intelligent control systems are associated with the wireless sensor networks. As a result, sensing and information processing in the sensor networks becomes more and more important. The increasing trend of more connected devices via wireless channels leads to the potential problem of lack of free and usable radio frequencies (as a national resource) and brings up the dilemma for allow an opportunistic spectrum usage. Special solutions are needed to handle that problem.

1.1 Cognitive Radio in Wireless Communications

Cognitive telecommunication systems are a relatively new interesting direction in telecommunication research. Traditionally the radio frequencies have been divided between the interested parties by licensing. The party who has a license to use a given frequency band has exclusive rights to the band and no one else can use this band. Nowadays we are reaching to the situation where the attractive frequency bands are full and there are no more frequencies available to license out new and innovative applications. This situation makes development and implementation of new radio-based services more difficult all over the world. Recent studies have shown, that the available licensed radio spectrum is becoming more occupied, while the assigned spectrum is significantly underutilized. The licensed users do not use their spectrum in all locations and all times and it is possible to utilize the available spectrum more fully and effectively. Cognitive radio [III00, HNZ09, BGG+13] is a technology that was proposed about 18 years ago by J. Mitola III to solve the problem [MM99]. Within this paradigm the radio equipment will search unused frequencies by itself and sense the spectrum area in terms of presence of licensed

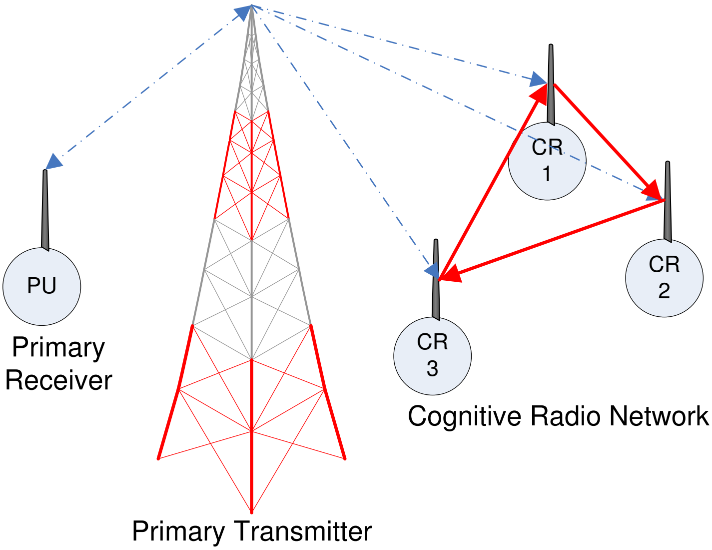

Figure 1.1: CR Principle

user. The proposed solution poses both technical and legal problems, which are currently dealt with. Cognitive radio is seen as a new promising technology and the research topic is providing interests to great amount of universities in spectrum sensing and signal detection, estimation, communication areas.

More specifically, spectrum utilization can be improved by allowing secondary (unlicensed) users to opportunistically access the licensed spectrum area when the primary user (PU) is not present. A cognitive radio (CR) technology is able to serve the secondary users for detecting and utilizing so called spectrum holes by sensing and adapting to the environment without causing harmful effects or interference to the licensed PUs. It is expected that CR systems are able to systematically detect the presence of a primary user (while the CR system usually does not have the *a priori* knowledge that the channel is free) by continuously sensing the spectrum area. If a PU signal is detected, the secondary user (SU) has to immediately stop operating in this specific frequency area and has to adapt and find new free spectrum area or channel for continuing operation. PUs may use different kind of modulations, transmission rates and powers, which makes the spectrum sensing more complicated. The CR network is illustrated in Fig. 1.1.

Since the active work-pattern of a PU is usually not known for the CR system, then adaptive signal processing methods could be used for spectrum sensing, which are able to learn and track the changes in the statistical properties of the underlying process.

One of the examples with Cognitive Radio technology is the usage of TV White Space. The unoccupied TV UHF band may be used for secondary services during time periods, when the primary TV stations are switched off [LZPH08, HNZ09]. Support for opportunistic spectrum access has for example been proposed initially for the LTE (4G) standard [OHMG12] and also for the 5G [LWM+17, HWWS14]. The topics related to cognitive radio technology are providing interests to world leading mobile access technology providers, including Ericsson.

In this thesis we investigate distributed cooperative detection algorithms that the radio equipment can use to determine whether a frequency is usable or not i.e. whether the primary user is using the frequency for its own purposes or not. A single cognitive user may not be in a good position to detect the presence of primary user with high probability because of the effects of radio propagation like fading and shadowing of radio waves. A more reliable decision can be obtained if several cognitive users work together sharing information. In the thesis we will investigate two cooperative detection techniques, that do not need any fusion center, which would be a single point of failure, but are rather similar to those used in adaptive filtering to share the information. The individual nodes will share the information directly with each other.

The aim is to develop algorithms usable for both individual and cooperative detection that can be used in cognitive radio networks to detect the presence of primary users. In this thesis we assume that there is only one PU signal present, however the current work can be logically extended also to the cases, were more PU signals are present, by updating the measurement signal model and by choosing or designing most optimal detector (module) for these specific cases.

1.2 Adaptive Distributed Signal Processing and Optimization

Several classical distributed detection methods have been proposed and studied in the literature and over decades [Var96]. Most of the classical solutions are however based on the "close to or ideal" *a priori* knowledge about the statistical properties of the observations and the detection hypotheses. In the CR application area, we have usually limited information about the PU signal and about the prior probabilities of the detection hypotheses. The CR system usually has limited information about transmission parameters, modes and functions of the PU system. Thus in the CR context we usually can not design an optimal detector in the sense of classical detection theory, since the parameters of the conditional distributions of the measurements are not ideally known (but have to be estimated, where an estimation error is always present). Also in CR context it is not that practical to limit the detection solutions with the assumption that the prior probabilities of the detection hypotheses are known and fixed over a period of time. Thus the classical detection methods based on the Bayesian approach are not that practical in CR context and we are aiming to use the Neyman-Pearson type of detectors in this thesis.

Adaptive filters [Hay02,Say08] have been used extensively in the systems, where the parameter to be estimated has a dynamic nature. Several applications in the literature use non-adaptive estimation methods (based on collected amount of samples) are used to estimate a parameter of interest. Adaptive (recursive) algorithms are however able to react to the changes in the statistical properties of the measurements "on line" and during the time when the recursive algorithm is kept running. In comparison, classical non-adaptive estimation methods have to be usually restarted, when the maximum amount of samples have been collected and when the value of the estimate has been calculated. This leads to the design issues, related to the size of measurement data windows for a specific application and there is a higher chance to miss the start moments of the transitions in the statistical processes of the measurements. Secondly adaptive algorithms usually do not require large amount of system memory, since only the data from the previous time instant should be stored into the memory. These mentioned aspects make the usage of adaptive estimation algorithms in the Cognitive Radio application context more practical.

Distributed adaptive estimation and detection schemes have been studied before in several papers [CS11a,CS10]. An optimal, matched filter based distributed detection scheme has been studied in [CS11b]. However in most cases we do not have any information about the waveform of the PU signals and hence we cannot design a matched filter based solution [CS11b]. LMS (Least Mean Square) based distributed estimation schemes have been investigated for example in [CS11a, CS10, CS11a]. In the thesis LMS (Least Mean Square) based adaptive estimation algorithms (which is a stohastic gradient based algorithm) are chosen due to the simplicity, robustness and good tracking abilities, compared to for example RLS (Recursive Least Squares) [LCS08].

Some recent developments in adaptation, learning, and optimization over networks have been published for example in [Say14, STC+13]. Diffusion Optimization Strategies [CS12,CS13] can be seen as a generalizations of Diffusion LMS estimation algorithm [CS10, CS11a].

1.3 Motivations and Objectives

We consider a scenario with a number of CR nodes in the network, which sense a spectrum area of interest. For simplification of the analysis, we additionally assume that the Gaussian noise floor is constant over the nodes. Several solutions have been proposed, that make use of a central processing unit to collect all the measurements over sensing period from all the nodes and make decisions about presence or absence of PU, for example [LZPH08,KLW09,WNK+10]. Instead of this, we expect that the measurements or estimates are exchanged between the CR nodes directly, without involvement of any central processing unit (fusion center). At every time instant new measurements or estimates from the neighbouring nodes become available. Thus CR nodes estimate the elements of the test statistics in their own location and make individual decisions about the detection hypotheses. Depending on the exact topology of the network, with such a solution communication in the network can be reduced (compared to solutions where nodes send their measurements to a fusion center, which sends the collected estimates back to the nodes after an iteration of the estimation process). This method saves energy, required for the data transmission of the single nodes (transmitters usually consume most of the power of a node). On the other hand this method enhances network failure resistance (in case of fusion center stops operating).

The above discussion naturally leads to the following main research topics which are addressed in this thesis:

1. Cooperative signal processing in Cognitive Radio Networks.

2. Distributed estimation and detection in Cognitive Radio, without using a FC.

3. Distributed Energy and Largest Eigenvalue detection in Cognitive Radio. Resulting detection performance analysis.

The main research concerns in the thesis are the following:

1. Removal of the central processing unit – a fusion center (FC) – from the domain of estimation and detection in the Cognitive Radio network. It is expected that CR network is able to estimate the test statistic of a detector and to detect the presence of the PU signal without the usage of any FC.

   The solutions in this thesis are based on the idea that distributed estimation schemes are used for designing distributed detection schemes, with no use of a FC. Thus the distributed detection schemes are based on the underlying distributed estimation strategies and topology in the Cognitive Radio Network.

2. We assume to have limited information about the type and properties of the PU signal and therefore an energy detection method becomes a usable solution. The energy detection method is implemented in a distributed way in CR network. Secondly, several type of correlation matrix based detection methods exist in the literature. We have chosen to study the Largest Eigenvalue detection method, which is similarly implemented in a distributed way in CR network.

3. Least Mean Square (LMS) type of adaptive estimation algorithms are based on the Stochastic gradient descent and the LMS estimates are modelled as random variables. Thus LMS type of algoritms are suitable for the estimation of a statistical moment based detection test statistics. Distributed Diffusion LMS algorithms have been already proposed and studied in the literature.

We adapt the Diffusion LMS algorithms to estimate the statistical moment based detection test statistics directly.

4. Since the PU signal is assumed to be slowly fading, then we design the usage of distributed adaptive estimation schemes so that approximately equal statistical properties of the estimates are achieved in every CR node in the network. In such a way an averaged detection performance in every CR node is achieved regardless of the actual channel conditions of each single node.

5. Usage of adaptive and recursive estimation schemes. We are interested in the online tracking ability of the statistical properties of the estimates to react to the changes in the presence of PU signal − i.e to the changes of the underlying detection hypothesis, over the iteration period of the distributed estimation algorithm.

6. As common in the area of statistical signal processing an extensive performance analysis for the proposed algorithms is performed. Since the detection performance of the proposed distributed detection schemes depends on the statistical properties of the underlying estimates, we propose to use a generic framework for studying the performance of the proposed estimation schemes in the CR network level. We focus on the analysis of the theoretical statistical moments of the estimates to study the resulting detection performance.

7. In the simulation sections we compare the theoretical findings with the results, obtained via the Monte-Carlo based computer simulations. A good match between theory and practice allows us to use computationally much faster theoretical calculations to evaluate the performance of the proposed algorithms in different use cases. We use mainly the probability of detection versus averaged SNR type of computer simulations to study the detection performance of the proposed algorithms, to evaluate the ranges when the detection methods fail to provide perfect detection results.

1.4 Thesis Outline and Contribution

This section provides an outline of the thesis with a brief summary of the material presented in each chapter. This thesis consists of 5 chapters, the summary of which are as follows.

Chapter 2

Chapter 2 provides background information. Here we will briefly discuss the concepts and tools that are needed to follow the rest of the thesis. We give a short summary of the theory of statistical signal processing in connection to the material in this thesis, where we discuss the basics of detection and estimation theory. We provide a generic introduction for the derivation of Diffusion LMS type of algorithms. Also we provide a short summary about the literature on Cognitive Radio.

Chapter 3

Chapters 3 and 4 discuss the main contributions of this thesis. Each chapter follows the structure of the corresponding published papers and thus is complete by itself - the reader does not need the content of previous or subsequent chapters to follow the material. However, the chapters themselves address problems and solutions which are partly related. Each chapter begins with a “Background” section, which gives the overall context to the discussion that follows and ends with a “Conclusion” section which summarizes the chapter along with the main concepts from that chapter.

Thus more specifically, chapter 3 addresses the distributed energy detection problem in Cognitive Radio networks. Often we have limited information about the signal received by the cognitive radio nodes and such signal flow can not be modelled as a deterministic process. Since the radio signals contain information when the PU signal is present, then it is often more suitable to model the PU signal also by a random process, in addition to the radio channel noise process. In such cases an energy detection becomes a usable solution. We are interested to remove a potential single point of error - a central processing unit from the cognitive radio network. Each CR node should be able to rely only on the communication between the neighbour CR nodes. We use distributed recursive estimation schemes to estimate the power of the received signal in a distributed way.

We propose the usage of distributed, diffusion least mean square (LMS) type of power estimation algorithms and three different static network topologies: Ring-Around, Combine And Adapt and Adapt and Combine are studied. We provide a generic framework for studying the detection performance of the proposed schemes by using the statistical properties of these distributed estimates. In case of the Ring-Around topology, a generic recursive signal power (statistical variance) estimation algorithm is proposed and more specific results about the moment estimation of the distributed estimates can be given. These results have been integrated into the same chapter. The theoretical findings are verified by MATLAB based simulations.

This chapter is based on the following 3 papers:

[A] A. Ainomäe, T. Trump and M. Bengtsson, “Distributed Recursive Energy Detection,” *IEEE Wireless Communications and Networking Conference WCNC 2014*, Istanbul, Turkey, Nov 2014, pp. 176-183.

[B] A. Ainomäe, T. Trump and M. Bengtsson, “CTA Diffusion Based Recursive Energy Detection,” *CSCS 14, WSEAS Latest Trends in Circuits, System Signal Processing and Automatic Control*, Salerno, Italy, Jun 2014, pp. 38-47.

[C] A. Ainomäe, T. Trump and M. Bengtsson, “Distributed diffusion LMS based energy detection,” *IEEE 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, St. Petersburg, Russia, Nov 2014, pp. 176-183.

Chapter 4

Chapter 4 deals with distributed correlation matrix (CM) based signal detection in Cognitive Radio network. The PU signal is assumed to be temporally correlated. Similarly as in previous chapter, in this chapter we study the usage of diffusion LMS based estimation strategies for estimating the elements of the correlation matrices, used for PU signal detection. Two static network topologies Combine and Adapt (CTA) and Adapt and Combine (ATC) are used in this chapter and we add few simulations with Consensus and FC based network topology for comparison. The estimation and detection solution does not rely on any central processing unit in the network. The estimation strategies and the section of performance analyses have been adapted and extended to deal with vector estimates and block-covariance matrices. Several correlation matrix based detection solutions have been proposed in the literature and in this research work we have chosen the Largest Eigenvalue based detection solution, where in case of Primary user signal exists in the network we assume, that the PU signal has a rank one correlation matrix. In order to obtain analytic results on the detection performance, the exact distribution of the CM estimates are approximated by a Wishart distribution, by matching the moments. The theoretical findings are similarly verified by MATLAB based simulations.

This chapter is based on the following 2 papers:

[D] A. Ainomäe, T. Trump, M. Bengtsson, "Distributed Largest Eigenvalue Detection," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, New Orleans, USA, March 2017.

[E] A. Ainomäe, M. Bengtsson, T. Trump, "Distributed Largest Eigenvalue Based Spectrum Sensing using Diffusion Adaptation," *Accepted to IEEE Transactions on Signal and Information Processing over Networks*, Sept 2017.

Chapter 5

Finally, Chapter 5 summarizes the author's topics in this thesis and lists possible directions for future research.

Contributions by the author and Copyright Notice

As specified in the Section 1.4, material presented in this thesis is based on the author's previous work which is published or submitted to conferences and journals held by or sponsored by IEEE and WSEAS publishers. They hold the copyright of the published papers and will hold the copyright of the recently accepted papers.

The contributions of the author of this thesis on the included papers are the outcome of the author′s own work, in collaboration with the co-authoring academical advisors Prof. Mats Bengtsson and Prof. Tõnu Trump. Most of the problem formulations and initial ideas for the papers were proposed by the advisors Prof. Tõnu Trump and Prof. Mats Bengtsson. The author of this thesis is the first author of the papers A to E and has been giving the substantial and the vast majority of the contributions, especially derivation and implementation of the proposed algorithms, regarding theoretical analysis, computer simulations and paper writing. The second and third authors were helpful with technical discussions and proofreading.

Chapter 2

Preliminaries

In this chapter, we will introduce some basic concepts that are essential to follow the rest of thesis.

2.1 Summary on Cognitive Radio

In this section we provide a brief summary about the aspects of Cognitive Radio Networks, which are essential in the context of the thesis. The section is based mainly on the material from [HNZ09], [GSMS09].

It was already briefly mentioned in Chapter 1, that since frequency spectrum is a limited resource for wireless communications, then it may become congested. To meet these growing demands, some national frequency regulation institutions, such as US Federal Communications Commission (FCC), has expanded the use of the unlicensed spectral band. However traditional wireless communications systems are not able to adaptively utilize the frequency bands. Many studies show that while some frequency bands (for example allocated statistically for some licensed users) in the spectrum are heavily used, other bands are largely unoccupied most of time. These potential spectrum holes result in the under-utilization of available frequency bands.

The concepts of software−defined radio and cognitive radio have been recently introduced to enhance the efficiency of frequency spectrum usage in next generation wireless and mobile computing systems. Cognitive ratio, which can be implemented through software−defined radio, is able to observe, learn, optimize, and intelligently adapt to achieve optimal frequency band usage.

Dynamic spectrum access (DSA) or opportunistic spectrum access (OSA) is the key approach in a cognitive radio network and has emerged as a new design paradigm for next generation wireless networks. Therefore also a new spectrum licensing paradigm needs to be initiated by the national frequency regulation institutions, for being more flexible in allowing unlicensed (or secondary) users to access the spectrum as long as the licensed (or primary) users are not interfered with. Such a way the utilization of the frequency spectrum could be improved. Development of dynamic spectrum access-based cognitive radio technology has to in general deal with technical and practical considerations as well as regulatory requirements.

The main frequency bands for CR are considered as follows

1. UHF band, typically 470-790 MHz,
2. Cellular bands, typically 800-900 MHz, 1.8-1-9 GHz, 2.1 GHz, 2.3 GHz, and 2.5 GHz,
3. Fixed wireless access bands, typically 2.5-3.5 GHz.

Main functions of CR to support DSA can be listed as follows [HNZ09]:

1. Periodical spectrum sensing, which can be centralized (FC based) or distributed, to determine if the frequency area of interest is free,
2. Spectrum analysis, to process the information from previous task, plan the spectrum access and optimize the transmission parameters,
3. Spectrum access, with the help of a cognitive medium access control (MAC) protocol,
4. Spectrum mobility, to change the operating frequency band of CR users.

Three major models of dynamic spectrum access are listed: common-use, shared-use, and exclusive-use models. In the first case the spectrum is open for access to all users. In the second case licensed users (i.e. PUs) are assigned to the frequency bands which are opportunistically accessed by the unlicensed users. In the latter case a PU can grant access of a particular frequency band to an unlicensed user for a certain period of time.

CR has to use a frequency area without causing interference to the PUs. There are three main approaches for opportunistic spectrum access [GSMS09]:

1. Spectrum Interweave,
2. Spectrum Overlay,
3. Spectrum Underlay.

The interweave paradigm of operation was the original motivation for the idea of CR. The requirement is that the CRs should not interfere with the communication between the already active PUs. Thus the CRs should be able to detect (sense), with very high probability, the primary user transmissions in the network. Once the CR successfully detects the PU transmissions, it can opportunistically communicate only if it is able to do so without harming the PU transmissions. This requires spectrum agility or the ability to transmit at different frequencies. The temporary space−time−frequency gap in the transmission of primary users is referred to as a spectrum hole or a white space. The overlay paradigm is more advanced. The CR needs to know the channel between the primary transmitter and the primary and secondary receivers as well as the channel between the secondary transmitter and the primary receiver. With the channel knowledge of both the primary and CRs, the CR can then choose appropriate transmission strategies so that the communication in the secondary network causes least interference to the primary network. In the underlay paradigm, the secondary transmitter keeps the interference levels below a certain threshold. The primary receiver sees a higher noise level if the primary and secondary transmission overlap in the same band. Possible methods include transmission power control, beam-forming and spread spectrum techniques.

Combining of these methods may be also considered. Although the overlay and interweave approaches are similar, in this thesis we focus on the detection methods, which follow interweave approach. The detectors are not aware about the channel gains of PU signal.

2.1.1 Spectrum Sensing in Cognitive Radio

In this section, we briefly focus on the spectrum sensing task of CR. The objective is to detect the presence of transmissions from licensed users. Three major types of spectrum sensing types are listed: non−cooperative, cooperative and interference−based sensing.

The usual model for signal detection is given based on the following idea

$$\begin{aligned} H_0 &: x(n) = v(n) \\ H_1 &: x(n) = \alpha s(n) + v(n), \end{aligned} \tag{2.1.1}$$

where $x_k(n)$ is the received signal of a CR user at time instant n, $s(n)$ is the transmitted signal of the PU, $v(n)$ is the additive white Gaussian noise (AWGN), and α is the channel constant (gain).

Three classical and one class of additional detection methods in *non-cooperative sensing* are for example:

1. Matched filter detection or coherent detection,
2. PU transmitter energy detection,
3. Cyclostationary feature detection,
4. Correlation Matrix based detection.

The matched filter is generally used to detect a signal by comparing a known signal (i.e. a template) with the received signal. A matched filter will maximize the received SNR for the measured signal [Kay98]. If the information of the signal from a licensed user is known, then a matched filter is an optimal detector in stationary Gaussian noise [SHT04]. Thus when a signal template is perfectly known, a matched filter requires only a small amount of time to operate. On the other hand, when this template is not available or is incorrect, the performance of spectrum sensing degrades significantly. Matched filter detection is suitable when the PU signal has a pilot, preambles, synchronization word or spreading codes, which can be used to construct the template for spectrum sensing.

Energy detection is the optimal method for spectrum sensing when the information from a PU (i.e signal type, pattern etc) is unavailable [SHT04]. The output signal from a bandpass filter is squared and integrated over the observation interval. A decision algorithm compares the integrator output with a threshold to decide whether a licensed user exists or not. In general, the energy detection performance deteriorates, when the SNR decreases. The Energy detection method is studied further in Chapter 3 of this thesis.

The PU signal has often a cyclostationarity (periodic) pattern, which can be used to detect the presence of a licensed user. A signal is cyclostationary (in the wide sense) if the autocorrelation is a periodic function. With such periodic pattern, the transmitted PU signal can be distinguished from noise, which is a wide-sense stationary signal without correlation. In general, cyclostationary detection can provide a more accurate sensing result and it is robust to variations in noise power. However, the detection is complex and requires long observation periods to obtain the sensing result.

A second large group of detectors for spectrum sensing are based on eigenvalue properties of an estimated correlation matrix [TW12, WTL14, ZL09]. When the PU signal exploits certain type of low rank correlation, then this feature can be used to detect the presence of a PU signal. Several CM based detectors have been proposed in the literature: the largest eigenvalue (LE) method, the volume based detector (VD), the covariance based detector (CAV), which have been studied for example in [HSQ14,HQXZ15] and [ZC09]. So called robust detectors do not require noise power value in the threshold calculation. Eigenvalue Arithmetic to Geometric Mean (AGM) [HFL+15], the Maximum to Minimum eigenvalue ratio (MME), the Energy to Minimum Eigenvalue ratio (EME) [ZL09], the Eigenvalue Moment ratio (EMR) [HFL+15], and the Hadamard [HXZ15] robust detectors have been proposed in the literature. The LE method is studied further in Chapter 4 of this thesis.

An unlicensed transmitter may not always be able to detect the signal from a licensed transmitter due to its geographic separation (a shadowing problem) and channel fading (a multipath fading problem). In *cooperative sensing*, spectrum sensing information from multiple CRs are exchanged among each other to detect the presence of a PU. The cooperative spectrum sensing is usually performed in a centralized or distributed manner. Obviously cooperative sensing will increase the communication and computation overhead compared with non-cooperative sensing. However in case of cooperative sensing, the detection probability can usually be significantly improved [LZ09]. In this thesis we assume that fully distributed CR nodes perform spectrum sensing and no central processing unit is used in estimation and detection domain.

We also mention, that in case of *Interference based sensing*, the noise/interference level (from all sources of signals) at the receiver of the primary user is measured.

This information is used by a CR to control the spectrum access (e.g. by computing expected interference level) without violating the interference temperature limit. Alternatively, an unlicensed transmitter may observe the feedback signal from a licensed receiver to gain knowledge on the interference level.

Finally we briefly list the potential application areas of CR [HNZ09], [GSMS09]:

1. Next Generation Wireless Networks, Machine-to-machine communications (IoT), Dynamic spectrum access in cellular systems.
2. Wireless broadband for distribution and backhaul, Data boost for mobile networks,
3. Coexistence of different wireless technologies, Cognitive digital home
4. Intelligent transportation system, Long range vehicle-to-vehicle networks,
5. eHealth services,
6. Emergency networks,
7. Military networks.

2.1.2 Common Research areas in cognitive radio

For an overview, we list some main CR research areas and aspects, which follow the function areas of CR:

1. Spectrum sensing,
2. Spectrum management,
3. Spectrum mobility,
4. Network layer and transport layer,
5. Cross-layer design for cognitive radio networks,
6. Artificial intelligence approach in cognitive radio.

By following the recently emphasised interests in the world-level scientific conferences of communication systems, such as IEEE GLOBECOM 2017 but also IEEE ICCASP 2017, IEEE WCNC 2017, we can add the following. In the research area of embedded (electronic) systems a continuing interest is on the design of (energy and failure) efficient hardware platform and architectures for testing and implementing the CR technology. On the other hand in the research area of applications and services of CR, the continuing interest are in the areas of cognitive networking in TV whitespaces, adaptation and integration with newest access technologies (incl.

massive MIMO and full-duplex). Also aspects related to the (cyber-) security and privacy in CR radio networks are gaining an interest.

Let us mention, that since the development of new generation 5G access technology is closely related to the IoT (Internet of Things) concept, then recently the research area of CR in the 5G/IoT technologies has gained increasing interest. It is expected that 5G will become the backbone for IoT devices by forming an ecosystem of so called smart devices. For example [MMB16, Chapters 4 and 2] give an overview about the challenges related to the implementation of IoT using CR capabilities in the future 5G Mobile Networks. As also initially planned for 4G, in 5G technology, the CR technology is expected to improve the handling of resources of the future smart environments - such as improving the utilization of available radio spectrum.

Since in this thesis we focus on the area of spectrum sensing, then we specify, that generic research issues can be categorized for example as follows:

1. Sensing interference limit,
2. Spectrum sensing in multiuser and multichannel networks,
3. Optimizing the period of spectrum sensing,
4. Spectrum management issues,

where obviously the research in this thesis is related to the second topic (and with the focus on the physical layer).

2.1.3 Standardization in cognitive radio

In this section, we give few comments about the standardization in CR area, based on [GSMS09].

In May 2004 US Federal Communications Commission (FCC) initiated the proposal to provide more efficient and effective use of the TV spectrum (i.e in the VHF and UHF band). As a result, IEEE 802.22 Working Group (WG) was formed to define a standardized air interface based on CRs. The IEEE 802.22 standard for Wireless Regional area Networks) requires that CR nodes sense the spectrum to detect the presence or absence of active primary transmitters. In November 2008, the FCC issued second report to adopt rules to allow unlicensed radio transmitters to operate in TV white spaces in order to make a significant amount of spectrum available for new and innovative products and services, including broadband data and other services for businesses and consumers. FCC expects that a database and active spectrum sensing is used by the solution. In September 2010, the FCC released third report that finalized the rules for using unused TV bands for unlicensed wireless devices, where mandatory sensing requirements were removed.

Some of the other IEEE standards related to white space networks are as follows.

1. The IEEE 802.11af WG for channel access and coexistence in TV White Spaces (TVWS).
2. The P1900 WG for developing supporting standards dealing with new technologies and techniques being developed for cognitive radio and advanced spectrum management.
3. The IEEE SCC41for of checking, whether reusing the IEEE 802 PHY/MAC is optimal for white space operation and to estimate how far the performance of the system could benefit from a tailored PHY/MAC system.
4. The IEEE 802.19 focuses on developing standards for coexistence between wireless standards of CR devices. The standard was formed to minimize the interference between different networks belonging to various wireless standards in the unlicensed band.

The International Telecommunication Union (ITU) has formed the following study groups that discuss cognitive radio networks.

1. ITU-R Study Group 1 on Spectrum Management, dynamic spectrum issues was covered by working part 1B.
2. ITU-R Study Group 5 on Terrestrial Services, working part 5A has described the potential application of cognitive radio systems in the land mobile service.
3. ITU-R Study Group 5, working party 5D, where the scope of this work is to consider the inclusion of CRS into the IMT family of technologies.

In Europe:

1. The European Communications Committee (ECC), has a special Task Group working on operation of cognitive radio systems in the white spaces of the UHF frequency band. The initial focus is on opportunistic use of radio spectrum in TV White Spaces.
2. The End−to−End Efficiency is a German Large Scale Integrating Project for integrating cognitive wireless systems in the Beyond 3G (B3G) world. The key objective of the E3 project is to design, develop, prototype, and showcase solutions to guarantee interoperability, flexibility, and scalability between existing legacy and future wireless systems.

2.2 Detection and Estimation Theory

We will start with some elements of detection and estimation theory. This overview section is written based mainly on materials from the [Kay93], [Kay98] and [Yaj17, Ch.2]. Detection theory deals with the problem of determining a particular hypothesis from the observation, $\mathbf{x}$. Typically a hypothesis maps to a particular phenomenon that is being detected. For example, in the context of a CR, we can formulate a hypothesis for whether a PU signal is present or not. If there are only two hypotheses, H_0 and H_1 for a phenomenon, then the detection problem reduces to a binary hypothesis test. For a binary hypothesis, the following types of errors can occur when deciding based on the observation:

- A type-1 error or false alarm, which occurs when the observation is decoded as H_1, for an H_0 event. Probability of false alarm, $P_{FA} = \Pr(H_1; H_0)$[1].
- A type-2 error or miss, which occurs when the observation is decoded as H_0, for an H_1 event. Probability of miss, $P_M = \Pr(H_0; H_1)$.

For an optimal design, both type-1 and type-2 errors cannot be reduced simultaneously. A typical approach is to fix the false alarm (type-1 error) and seek an optimal detector to minimize the type-2 error. Note that minimizing the type-2 error is the same as maximizing the detection probability, $P_D = (1 - \Pr(H_0; H_1)) = \Pr(H_1; H_1)$. This setup is called the Neyman-Pearson (NP) approach to hypothesis testing. We can formalize this into a theorem as follows:

Theorem 2.2.1. *For a given false alarm, $P_{FA} = \alpha$, to maximize, P_D, decide toward H_1 if,*

$$L(\mathbf{x}) = \frac{p(\mathbf{x}; H_1)}{p(\mathbf{x}; H_0)} > \gamma, \tag{2.2.1}$$

where the threshold, γ, is obtained from

$$P_{FA} = \int_{\mathbf{x}:L(\mathbf{x})>\gamma} p(\mathbf{x}; H_0) d\mathbf{x} = \alpha. \tag{2.2.2}$$

Equation (2.2.1) is called the likelihood ratio test [Kay98]. Let us note that the formula for P_D is obviously given as

$$P_D = \int_{\mathbf{x}:L(\mathbf{x})<\gamma} p(\mathbf{x}; H_1) d\mathbf{x}. \tag{2.2.3}$$

In practice and given the specific signal model, the conditional probability density functions $p(\mathbf{x}; H_1)$ and $p(\mathbf{x}; H_1)$ of the observation variable x are specified. By following the standard derivation procedure, then usually all the constant variables in (2.2.1) are moved on the right side of the inequality and the observation data dependant variables on the left hand side. In general the detection formula can be given as follows

$$\begin{aligned} H_0 &: \mathbf{T}_x < \gamma, \\ H_1 &: \mathbf{T}_x \geq \gamma, \end{aligned} \tag{2.2.4}$$

[1] We define $\Pr(H_i; H_j)$ as the probability of choosing hypothesis H_i when H_j has occurred.

where after the mentioned steps the left hand side of the likelihood ratio is made equal to the variable $\mathbf{T}_x$, which is called a test statistics of the detector. The exact or approximate conditional probability density functions are assigned for the variable $\mathbf{T}_x$ and as mentioned above. Througout the thesis, the threshold γ is determined based on the desired P_{FA} value. Often the detection performance of a NP detector is studied with the help of P_D versus P_{FA} graphs, called as Receiver Operation Charateristics (ROC) [Kay98, Chapter 3.4].

The details for the Energy and Largest Eigenvalue Detectors are given in the corresponding sections of Chapters 3 and 4 respectively. In this thesis swe use the P_D versus the network average SNR graphs to study the areas there the detection method fails to provide perfect detection results.

The Estimation theory deals with arriving at a quantitative conclusion about a parameter, $\boldsymbol{\theta}$, from the observation, $\mathbf{x}$. An example of this is estimating the power of the PU signal (which is modelled as a CSCG process) in CR network from a function of received PU signal samples. The joint probability distribution function, $p(\boldsymbol{\theta}, \mathbf{x})$, denotes the complete statistical description of the parameters and observations. The parameter, $\boldsymbol{\theta}$ can be random and unknown. However, in certain estimation problems, $\boldsymbol{\theta}$, can be deterministic. Under these conditions, good estimators can be designed by mathematically modelling the observation $\mathbf{x}$, through the parametrized, PDF, $p(\mathbf{x}; \boldsymbol{\theta})$.

Typical estimation methods depend on the model assumptions. In this thesis we deal mainly with the mean and variance estimation tasks. The details are described in Section 3 and 4 respectively.

Let us note that in case of a PU signal detection problem, the usage of Bayes approach, both in detection [Kay98] and also in estimation [Kay93] domain, is rather impractical, since usually the CR system does not obtain sufficiently accurate and *a priori* data about the (longer time) statistical behaviour of the PU signal(s) and thus about the parameters of the distributions of the corresponding random processes. It is more practical to view the PU behaviour as a dynamic process, where the statistical parameters of interest may change inexplicably during the observation time. Thus we rather need to look for the adaptive estimation solutions to implement the detectors of interest.

2.3 Adaptive Distributed Signal Processing and Optimization

An adaptive filter is a system with a linear filter that has a transfer function controlled by variable parameters and a means to adjust those parameters according to an optimization algorithm [Say08,Hay02]. Usually the adaptive filters are digital filters and are suitable for the applications where some parameters of the desired processing operation are not known in advance or are changing over the time instant.

Stochastic optimization methods are optimization methods that generate and use random variables. For stochastic problems, the random variables appear in the formulation of the optimization problem, which involve for example random objective functions. Stochastic gradient descent is a stochastic approximation of the gradient descent optimization method for minimizing an objective function − i.e by finding a minima or maxima by iteration. A popular stochastic gradient descent algorithm is the least mean squares (LMS) adaptive filter.

Thus the concepts of adaptive filtering and stochastic optimization are connected. In both cases usually a parameter of interest is found from the realizations of random inputs variables iteratively by solving an optimization problem with the minima search.

In resent years the research area of distributed optimization has gained increasing interest [Say14, Say12]. Distributed estimation algorithms are useful in several contexts, including wireless and sensor networks, where scalability, robustness, and low power consumption are desirable. Since diffusion cooperation schemes (such as diffusion LMS) have been shown to provide good performance, robustness to node and link failure and are amenable to distributed implementations [CS10], then in this thesis we have used diffusion LMS type of algorithms for designing and implementing the distributed Energy and Largest Eigenvalue detection solutions.

2.3.1 Diffusion LMS Algorithm

In this overview section we briefly describe the idea and the derivation steps of the distributed, Diffusion Least Mean Square type of algorithm and in general form, by summarising the material from [CS10]. This section provides some brief background info for the reader to follow the re−derivation and implementation steps of the diffusion LMS type of algorithms in Chapters 3 and 4.

Distributed Estimation Problem Formulation

Let us assume we have K nodes in CR network. Let $\mathcal{N}_k$ denote the neighborhood group of node $k \in K$, i.e $\mathcal{N}_k$ defines the set of nodes l which can send data unidirectionally the node k. In general, at time instant n, every node k receives:

1. a scalar measurement $d_k(n)$ and a $1 \times M$ row regression vector $\mathbf{u}_{k,n}$ or
2. a $M \times 1$ vector measurement $\mathbf{d}_k(n)$ and when the row regression vector $\mathbf{u}_{k,n}$ is neglected from the derivations.

$d_k(n), \mathbf{d}_k(n), \mathbf{u}_k(n)$ are realizations of corresponding complex random processes. On page 24 we explain that in this thesis we adapt and apply the theory of diffusion LMS for two different measurement and estimation dimension sets. In the first case every node k, using data set $\{d_k(n), \mathbf{u}_k(n)\}$ estimates an optimal parameter p^o. In the second case an optimal $M \times 1$ vector $\mathbf{p}^o$ is estimated based on the set $\{\mathbf{d}_k(n)\}$. Thus for the generic notation in this overview section, we use boldface notation

$\mathbf{d}_k(n)$ for the measurement parameter and $\mathbf{p}^o$ for the optimal vector respectively and show the row regression parameter $\mathbf{u}_{k,n}$ in the derivations.

Global Optimization

We seek the $M \times 1$ optimal linear estimator $\mathbf{p}^o$, that minimizes the following **global cost** function

$$J^{glob}(\mathbf{p}) \triangleq \sum_{k=1}^{K} \mathrm{E}\,|\mathbf{d}_k(n) - \mathbf{u}_{k,n}\mathbf{p}|^2. \tag{2.3.1}$$

In case of the so called "desired process" $\mathbf{d}_k(n)$ and the "regressor process" $\mathbf{u}_{k,n}$ are Wide Sense Stationary, then the optimal solution is given as

$$\mathbf{p}^o = \left(\sum_{k=1}^{K} \mathbf{R}_{u,k}\right)^{-1} \left(\sum_{k=1}^{K} \mathbf{R}_{du,k}\right), \tag{2.3.2}$$

where $\mathbf{R}_{u,k} = \mathrm{E}\left[\mathbf{u}_{k,n}^*\mathbf{u}_{k,n}\right]$ and $\mathbf{R}_{du,k} = \mathrm{E}\left[\mathbf{d}_k(n)\mathbf{u}_{k,n}^*\right]$ are the corresponding covariance matrices.

Steepest Descent Solution

For the minimization of the global cost function, standard iterative Steepest-Descent algorithm can be used and we have

$$\mathbf{p}_n = \mathbf{p}_{n-1} - \mu \left[\bigtriangledown_w J^{glob}(\mathbf{p}_{n-1})\right]^*, \tag{2.3.3}$$

where scalar step size parameter is $\mu > 0$ and $\mathbf{p}$ is the estimate of $\mathbf{p}^o$, at time iteration i. Complex gradient is given as follows

$$\left[\bigtriangledown_{\mathbf{p}} J^{glob}(\mathbf{p}_{n-1})\right]^* = \sum_{k=1}^{K} \left(\mathbf{R}_{u,k}\mathbf{p} - \mathbf{R}_{du,k}\right), \tag{2.3.4}$$

and we get the steepest descent recursion as

$$\mathbf{p}_n = \mathbf{p}_{n-1} - \mu \sum_{k=1}^{K} \left(\mathbf{R}_{du,k} - \mathbf{R}_{u,k}\mathbf{p}_{n-1}\right). \tag{2.3.5}$$

Since usually the second order moments in (2.3.5) are not known *a-priori*, then the following approximations can be used instead: $\mathbf{R}_{u,k} \approx \mathbf{u}_{k,n}^*\mathbf{u}_{k,n}$ and $\mathbf{R}_{du,k} \approx \mathbf{d}_k(n)\mathbf{u}_{k,n}^*$. As a result, we get a non−distributed Global [2] LMS type of algorithm

$$\mathbf{p}_n = \mathbf{p}_{n-1} - \mu \sum_{k=1}^{K} \mathbf{u}_{k,n}^* \left(\mathbf{d}_k(n) - \mathbf{u}_{k,n}\mathbf{p}_{n-1}\right). \tag{2.3.6}$$

[2] The term Global means that the algorithm requires data from all the nodes in the network.

Local Optimization

Introduce a matrix $\mathbf{C}$ with elements $\{c_{l,k}\}$, where the element $c_{l,k}$ defines if observation from node l is available for the node k. $\mathbf{C}$ is usually considered to be doubly-stochastic $K \times K$ non-negative real matrix with entries $c_{l,k}$ and $c_{l,k} = 0$ if $l \notin \mathcal{N}_k$ and thus obviously $\mathbf{C1} = \mathbf{1}$, $\mathbf{1}^T\mathbf{C} = \mathbf{1}^C$. The **local cost** at node k is given as

$$J_k^{loc}(\mathbf{p}) = c_{l,k}\,\mathrm{E}\,|\mathbf{d}_l(n) - \mathbf{u}_{l,n}\mathbf{p}|^2. \tag{2.3.7}$$

The optimal solution can therefore be updated

$$\mathbf{p}_k^{loc} = \left(\sum_{l\in\mathcal{N}_k} c_{l,k}\mathbf{R}_{u,l}\right)^{-1}\left(\sum_{l\in\mathcal{N}_k} c_{l,k}\mathbf{R}_{du,l}\right). \tag{2.3.8}$$

Define additionally the matrix $\Gamma_k \triangleq \sum_{l\in\mathcal{N}_k} c_{l,k}\mathbf{R}_{u,l}$. By completing the squares, we get that J_k^{loc} can be alternatively rewritten in terms of $\mathbf{p}_k^{loc}$ as

$$J_k^{loc}(\mathbf{p}) = \|\mathbf{p} - \mathbf{p}_k^{loc}\|^2_{\Gamma_k} + \mathbf{MMSE}, \tag{2.3.9}$$

where the **MMSE** is a constant part. By using the matrix $\mathbf{C}$ then minimizing of the global cost $J^{glob}(\mathbf{p})$ is equivalent to minimizing of the following cost function for any $k \in K$

$$J^{glob}(\mathbf{p}) = \sum_{l=1}^{K} J_l^{loc}(\mathbf{p}) = J_k^{loc}(\mathbf{p}) + \sum_{l\neq k}^{K} J_l^{loc}(\mathbf{p}) \tag{2.3.10}$$

$$J^{glob}(\mathbf{p}) = \sum_{l\in\mathcal{N}_k} c_{l,k}\,\mathrm{E}\,|\mathbf{d}_l(n) - \mathbf{u}_{l,n}\mathbf{p}|^2 + \sum_{l\neq k}^{K} \|\mathbf{p} - \mathbf{p}_k^{loc}\|^2_{\Gamma_l} \tag{2.3.11}$$

We have now an alternative global cost representation in terms of local estimates $\left\{\mathbf{p}_k^{loc}\right\}$.

MSE Minimization

Minimization of $J^{glob}(\mathbf{p})$ on every node k, still requires access to the global information $\left\{\mathbf{p}_l^{loc}\right\}$ and matrices Γ_l in the other nodes in the network. A fully distributed solution is derived at next and this is based on the diffusion LMS strategy.

Let us replace $\mathbf{\Gamma}_l$ with $\mathbf{\Gamma}_l = b_{l,k}\mathbf{I}_M$, where $\mathbf{I}_M$ is $M \times M$, $b_{l,k} = 0$ if $l \notin \mathcal{N}_k$, $\mathbf{1}^T\mathbf{B} = \mathbf{1}^T$. Let us introduce a new $K \times K$ matrix $\mathbf{B}$. Also we replace $\mathbf{p}_k^{loc}$ with the intermediate estimate $\boldsymbol{\psi}_l$ at node l. Then the following approximation of J^{glob} is proposed and so that each node k can minimize modified cost as

$$J_k^{dist}(\mathbf{p}) = \sum_{l\in\mathcal{N}_k} c_{l,k}\,\mathrm{E}\,\|\mathbf{d}_l(n) - \mathbf{u}_{l,n}\mathbf{p}\|^2 + \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}\|\mathbf{p} - \boldsymbol{\psi}_l\|^2 \tag{2.3.12}$$

The complex gradient is given as:

$$\left[\bigtriangledown_{\mathbf{p}} J_k^{dist}(\mathbf{p}_{n-1})\right]^* = \sum_{l\in\mathcal{N}_k} c_{l,k}\left(\mathbf{R}_{u,l}\mathbf{p} - \mathbf{R}_{du,l}\right) + \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}\left(\mathbf{p} - \boldsymbol{\psi}_l\right). \quad (2.3.13)$$

We can use $J_k^{dist}(\mathbf{p})$ to obtain the recursion for the estimate of $\mathbf{p}$ at node k in two steps:

$$\begin{aligned}\boldsymbol{\psi}_{k,n} &= \mathbf{p}_{k,n-1} + \mu_k \sum_{l\in\mathcal{N}_k} c_{l,k}\big(\mathbf{R}_{du,l} - \mathbf{R}_{u,l}\mathbf{p}\mathbf{p}_{k,n-1}\big)\\ \mathbf{p}_{k,n} &= \boldsymbol{\psi}_{k,n} + \nu_k \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}\big(\boldsymbol{\psi}_l - \mathbf{p}_{k,n-1}\big).\end{aligned} \quad (2.3.14)$$

In the second equation two replacements are performed: $\boldsymbol{\psi}_l$ is replaced by the intermediate estimate $\boldsymbol{\psi}_{l,n}$, available at node l, at time n, and secondly $\mathbf{p}_{k,n-1}$ is replaced by $\boldsymbol{\psi}_{k,n}$. As a result we get

$$\begin{aligned}\boldsymbol{\psi}_{k,n} &= \mathbf{p}_{k,n-1} + \mu_k \sum_{l\in\mathcal{N}_k} c_{l,k}\big(\mathbf{R}_{du,l} - \mathbf{R}_{u,l}\mathbf{p}_{k,n-1}\big)\\ \mathbf{p}_{k,n} &= \boldsymbol{\psi}_{k,n} + \nu_k \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}\big(\boldsymbol{\psi}_{l,n} - \boldsymbol{\psi}_{k,n}\big).\end{aligned} \quad (2.3.15)$$

The second recursion can be rearranged again. First recall that

$$\mathbf{p}_{k,n} = (1 - \nu_k + \nu_k b_{k,k})\boldsymbol{\psi}_{k,n} + \nu_k \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}. \quad (2.3.16)$$

Let us define $K \times K$ matrix left stochastic $\mathbf{A}$, which elements are the coefficients $a_{k,k} = (1-\nu_k+\nu_k b_{k,k})$ and $a_{l,k} = (\nu_k b_{l,k})$ for $l \neq k$. We get the following recursion

$$\begin{aligned}\boldsymbol{\psi}_{k,n} &= \mathbf{p}_{k,n-1} + \mu_k \sum_{l\in\mathcal{N}_k} c_{l,k}\big(\mathbf{R}_{du,l} - \mathbf{R}_{u,l}\mathbf{p}_{k,n-1}\big)\\ \mathbf{p}_{k,n} &= \sum_{l\in\mathcal{N}_k} a_{l,k}\boldsymbol{\psi}_{l,n}.\end{aligned} \quad (2.3.17)$$

Let us note that $c_{l,k} = a_{l,k} = 0$ if $l \notin K$, $\mathbf{1}^T C = \mathbf{1}^T$, $C\mathbf{1} = \mathbf{1}$, and obviously $\mathbf{1}^T A = \mathbf{1}^T$.

ATC and CTA Diffusion LMS algorithms

Next we summarise the **Adapt and Combine** (ATC) and **Combine and Adapt** (CTA) type of Diffusion LMS algorithms, by inserting the approximations of the covariance matrices.

ATC Diffusion LMS

Init: $\mathbf{p}_{k,0} = 0$ for all $k \in K$. Given the non-negative real coefficients $\{c_{l,k}, a_{l,k}\}$ for each time $n \geq 0$ and for all nodes k:

$$\begin{cases} \boldsymbol{\psi}_{k,n} = \mathbf{p}_{k,n-1} + \mu_k \sum_{l \in \mathcal{N}_k} c_{l,k} \mathbf{u}_{l,n}^{*} \left(\mathbf{d}_l(n) - \mathbf{u}_{l,n} \mathbf{p}_{k,n-1} \right), & \text{(incremental step)}, \\ \mathbf{p}_{k,n} = \sum_{l \in \mathcal{N}_k} a_{l,n} \boldsymbol{\psi}_{l,n} & \text{(diffusion step)}. \end{cases} \tag{2.3.18}$$

CTA Diffusion LMS

Init: $\mathbf{p}_{k,0} = 0$ for all l. Given the non-negative real coefficients $\{c_{l,k}, a_{l,k}\}$ for each time $n \geq 0$ and for all nodes $k \in K$:

$$\begin{cases} \boldsymbol{\psi}_{k,n-1} = \sum_{l \in \mathcal{N}_k} a_{l,k} \mathbf{p}_{l,n-1} & \text{(diffusion step)}, \\ \mathbf{p}_{k,n} = \boldsymbol{\psi}_{k,n-1} + \mu_k \sum_{l \in N_k} c_{l,k} \mathbf{u}_{l,n}^{*} \left(\mathbf{d}_l(n) - \mathbf{u}_{l,n} \boldsymbol{\psi}_{k,n-1} \right), & \text{(incremental step)}. \end{cases} \tag{2.3.19}$$

We note that detailed performance analysis of the Diffusion LMS algorithms is performed in [CS10] but in the estimation domain only and based on the estimation error recursions.

Comments on the implementation and usage in the CR context

In Chapters 3 and 4 we use the Diffusion LMS algorithm derivation framework for deriving a diffusion LMS based scalar (power) estimation solution for the distributed Energy detection solution and a diffusion LMS based vector (vectorized correlation matrix) estimation solution for the distributed Largest Eigenvalue detection solution. For deriving these latter estimation algorithms, we need to introduce small modifications in the standard derivation flow of the Diffusion LMS algorithms.

The considerations are the following.

1. Depending on the application of an adaptive filter [Hay02, Chapter 1.7], the regressor variable $\mathbf{u}_{k,n}$ can be seen as a variable, which can contain some *a priori* information for the estimation process. In a practical PU signal detection task a CR system usually can not use *a priori* data, which can be incorporated in the estimation process of the elements of test statistics – i.e the signal sequence of the PU user for implementing a matched filter detection solution. For the Energy and Largest Eigenvalue detection solutions, proposed in this thesis, the regressor variable is expendable (i.e $\mathbf{u}_k(n) = 1$ constantly) and thus can be excluded from the derivations. The secondary statistics becomes then $\mathbf{R}_{u,k} = 1$ and $\mathbf{R}_{du,k} = \mathrm{E}\left[\mathbf{d}_k(n)\right]$. Thus in our solutions the "desired" variable $\mathbf{d}_k(n)$ is connected with the observations for the estimation process.

2. Due to the previous point and for the power estimation algorithm in Chapter 3, the $\mathbf{p}^o$ and $\mathbf{d}_k(n)$ are both selected as scalars and the derivation of diffusion LMS type of algorithm can be slightly simplified. These details are shown in Chapter 3.

3. For the vector estimation algorithm in Chapter 4, the variables $\mathbf{d}_k(n)$ and $\mathbf{p}^o$ are taken as a $M \times 1$ vectors. The derivation of diffusion LMS type algorithm is slightly modified and these details are shown in Chapter 4.

4. In this thesis we do not proceed with the performance analysis of the Diffusion LMS type algoritm, based on the estimation error measures. Instead we are interested in the analysis of the statistical moments of the estimates directly, to proceed with the analysis of the detection performance of the proposed distributed detection solutions.

Thus in Chapters 3 and 4 we skip some of the standard derivation steps and focus on the differences from the standard derivation flow of diffusion LMS type of algorithms.

Chapter 3

Distributed Diffusion LMS based Energy Detection

CR systems need to detect the presence of a primary user by continuously sensing the spectrum area of interest. Radiowave propagation effects like fading and shadowing often complicate sensing of spectrum holes because the PU signal can be weak in a particular area. Cooperative spectrum sensing is seen as a prospective solution to enhance the detection of PU signals. This chapter studies distributed spectrum sensing in a cognitive radio context based on the results in [ATB14b] and [ATB14a]. We investigate distributed energy detection schemes without using any fusion center. Due to reduced communication such a topology is more energy efficient. We propose the usage of distributed, diffusion least mean square (LMS) type of power estimation algorithms. In this chapter an Adapt and Combine (ATC) diffusion based power estimation scheme is proposed and the performance is compared with the Combine and Adapt (CTA) and ring-around schemes in a common framework. Additionally we show in this chapter also the results from the first paper [ATB14c] for a recursive and distributed power estimation scheme with a ring around topology, which does not necessarily have to be related with the Diffusion LMS context. In this case specific theoretical results for the performance analysis of that algorithm can be given. The power to be estimated for the energy detection is a scalar quantity. The PU signal is assumed to be slowly fading. We analyse the resulting energy detection performance and verify the theoretical findings through simulations.

3.1 Background

The cognitive radio (CR) system is dynamic. Often in practice the statistical information (for example conditional probability density of observations, prior probabilities of detection hypotheses, longer time statistical behaviour of primary user (PU)) is not available *a priori* for constructing a PU signal detection solution. The properties of the test statistics (for making a detection decision) may change in time.

In cognitive radio context we would like to avoid interference to the PU user and find free spectrum opportunities as fast as possible. On-line distributed network learning methods are able to learn the statistical information based on observations received by the nodes in the network. These methods can react to possible changes in the properties of estimated statistics in real time.

Several proposed distributed spectrum sensing solutions make use of a central fusion center [HNZ09], [LZPH08], [KLW09], [WNK+10]. A fusion center is however seen as a single point of failure in the network since a malfunction in this unit affects the performance of the whole distributed solution. We propose a power estimation solution, where the available power estimates (and measurements) are fused in cognitive radio network nodes, to allow all nodes to make detection decisions based on data from the neighbour nodes and without involvement of any central processing unit. Such a solution enhances network failure resistance (at the cost of slightly increased information overhead in the network).

Several distributed adaptive estimation and detection schemes have been studied in the past. Least mean square (LMS) and recursive least squares (RLS) based estimation schemes are analysed for example in [CS11a], [CS10], [LS08], [LCS08] and consensus based schemes in [XBL06], [SMG09], [XBL05], [DKM+10]. Optimal, matched filter distributed detection, based on diffusion type LMS and RLS estimation schemes, was studied in [CS11b]. Here, we make the assumption that the CR network does not have any prior information about the waveform of the PU signal in the secondary nodes and hence we cannot design a matched filter. Therefore energy detection becomes a practical solution.

A ring network topology for distributed energy detection without a fusion centre has been suggested in [KGC11]. In [ATB14c] we proposed and analysed an estimation based recursive calculation of the test statistics for the energy detectors in cognitive radio network with ring topology. The test statistic in form of a converged power estimate is the soft information used for making the detection decision at every node. Ring networks are however sensitive to link failures. Combine and Adapt (CTA) diffusion based recursive calculation of the test statistics for the energy detectors was proposed and studied in [ATB14a]. In this chapter we focus mainly on the analysis of the Adapt and Combine (ATC) version of diffusion LMS type of received power estimation algorithm. The performance of the ATC diffusion based distributed power estimator is compared with the previously proposed CTA [ATB14a] and ring [ATB14c] schemes to complete the analysis. The resulting energy detection performance is studied and is dependent on the performance of the used distributed recursive power estimation algorithm.

We organize the remainder of the chapter as follows. In Section 3.2 we review the system model and the basics of energy detection. We derive an ATC type received signal power estimation algorithm based on diffusion LMS strategy and summarize the CTA based version. In Section 3.3 we analyse the performance of the proposed distributed power estimation algorithm (using a common model) and the resulting energy detection. In Section 3.4 we present our simulations results.

3.2 Distributed power estimation and detection

According to classical detection theory, an energy detector can be used for detecting random signals in additive noise. For energy detection in a cognitive radio context, the type of PU signal can be completely unknown. During a sensing time t, an energy detector (ED) receives N samples of a signal $x(n)$ from a specific frequency band [LZPH08]. The average energy of the received data samples is the test statistic $T(x)$ of the ED, which compares $T(x)$ to a predefined threshold γ and decides which of the hypotheses H_0 or H_1 is more likely.

We assume the following signal model at node k:

$$\begin{aligned} &H_0 : E[|x_k(n)|^2] = E[|v_k(n)|^2] \\ &H_1 : E[|x_k(n)|^2] = E[|\alpha_k|^2|s(n)|^2] + E[|v_k(n)|^2], \end{aligned} \tag{3.2.1}$$

where $k = 1, 2, ..., K$ is the node number and $n = 1, 2, ...N$ is the sample index. $v_k(n)$ is independent and identically distributed ($i.i.d$) circularly symmetric complex Gaussian (CSCG) noise with zero mean and variance $E[|v_k(n)|^2] = \sigma^2_{v,k}$, i.e. $v(n) \sim \mathcal{C}N(0, \sigma^2_{v,k})$. The power of the emitted PU signal $s(n)$ is denoted as $E[|s(n)|^2] = S$, under H_1. The primary signal $s(n)$ and the noise $v_k(n)$ are assumed to be statistically independent. The PU signal passes through a slowly fading channel with gain $\alpha_k(n)$. The gain α_k is considered to be constant. Note, that for implementing the energy detector, only the noise variance is needed to determine the detection threshold γ, therefore estimates of the channel gains are not required in practical implementations. Noise power estimation is not considered in this research work. In this chapter we make the following assumptions:

- (AS 1) The $x(n)$ is sensed by K nodes in the CR network.
- (AS 2) The additive noise $v_k(n)$ is uncorrelated in time and space and has the same power level over all the nodes in the CR network.
- (AS 3) The number of performed iterations N is large enough.
- (AS 4) The links between the CR nodes are ideal and not capacity restricted (no need to quantize the soft information).

In the literature on distributed detection, for example in [Var96], a fusion center, which collects all the local soft information, hard or soft binary decisions from the sensors, is often used in distributed detection networks. Similarly a central processing unit has been used in distributed estimation schemes, see e.g. [CS11a]. However, such a central processing unit can potentially be a single point of failure in the detection system. Secondly it may require frequent data exchange between the nodes and the centre and thus drain system energy resources, since usually most of the energy is spent for powering up the transmitter to exchange the data with neighbour nodes.

A distributed and recursive estimation scheme is one of the possible solutions for removing the central processing unit from the system and thus the network is able to calculate the global estimates based on the local observations collected by the CR nodes. Then based on the estimated test statistic, the detector at each CR node can locally make its own decision if the PU signal is present or not. We denote the power estimate at node k and at iteration n as $\hat{p}_k(n)$. The network topology is assumed to be fixed over the sensing time. We consider a linear, fixed combination of neighbour estimates and measurements at every node k.

Next we shortly review the global model for estimating the received signal power in cooperative manner (as proposed in [ATB14a]). Then we derive an ATC type power estimation algorithm, where the nodes can observe the measurements and share the estimates (and measurements) only with their neighbour nodes, according a to predefined network topology. Finally we propose a data exchange and combination strategy for ATC diffusion algorithm.

3.2.1 Global estimation

According to model (3.2.1), the power of the PU signal is attenuated at every node k. The locally estimated power varies between nodes k. Therefore if the channel gain at node k is low, the resulting energy detection performance is low. The result is opposite, when the node has a good channel gain. When nodes cooperate to estimate a common parameter P^o, the resulting detection performance will improve. As in [ATB14a] we recommend the following form of p^o

$$p^o = \frac{1}{K}\sum_{k=1}^{K} \mathrm{E}\left[|x_k(n)|^2\right] = S\frac{1}{K}\sum_{k=1} |\alpha_k|^2 + \sigma_v^2. \tag{3.2.2}$$

The p^o is the average of the received power across the nodes $k \in K$ in the network. The second equation in (3.2.2) follows from the signal model (3.2.1) if the PU signal is present and from the assumption AS 2. When we have sufficient number of nodes in the CR network, the effect of varying channel gains is averaged over nodes $k \in K$.

The corresponding global cost function is given as:

$$J^{glob}(p) = \sum_{k=1}^{K} \mathrm{E}\left[|x_k(n)|^2 - p\right]^2, \tag{3.2.3}$$

where we have used the form of global cost as proposed in [STC+13], [CS11b], [CS11a]. Minimization of the mean square error across the network (3.2.3) with respect to P results in the optimal solution, which is given by (4.3.5).

3.2.2 Distributed ATC Diffusion LMS estimation

Suppose that K nodes in the CR network are interested in estimating the scalar parameter p^0 in a distributed manner, where nodes rely only on the information, that is available to them. Depending on network topology, nodes are connected only to selected neighbour nodes and do not have access to any global data. The global cost (3.2.3) needs to be approximated in a distributed manner. The derivation of the ATC diffusion power estimation algorithm follows the ideas in [Say12], [CS10].

Let $\mathcal{N}_k$ denote the neighbourhood group of node $k \in K$, i.e $\mathcal{N}_k$ consists of nodes l which can communicate with node k. We assume that the network is connected and the connection between nodes l and k is unidirectional.

Let us define $K \times K$ doubly stochastic matrix $\mathbf{C}$ containing non-negative elements $c_{l,k}$ and $c_{l,k} = 0$ if $l \neq \mathcal{N}_k$ (i.e when data from node l is not available for node k). Let us note that for a doubly stochastic matric $\mathbf{C}$ it holds that $\mathbf{C1} = \mathbf{1}$ and $\mathbf{1}^T\mathbf{C} = \mathbf{1}^T$. The local cost and the corresponding local optimal solution in the neighbourhood of node k can be expressed with the help of coefficients $c_{l,k}$ as follows

$$J_k^{loc}(p) = \sum_{l \in \mathcal{N}_k} c_{l,k} \, \mathrm{E}\left[|x_l(n)|^2 - p\right]^2, \tag{3.2.4}$$

$$p_k^{loc} = \sum_{l \in \mathcal{N}_k} c_{l,k} \, \mathrm{E}\left[|x_l(n)|^2\right]. \tag{3.2.5}$$

The global cost can be fractioned into the local cost of over the neighbourhood of node k and local costs over the neighbourhood of other nodes. Using the completion of squares argument [STC+13] to relate variable P and local optimal solution P_l^{loc}, secondly ignoring the mmse part which is not dependant on p, the global cost function can be expressed as follows

$$J^{glob'}(p) = \sum_{l \in N_k} c_{l,k} \, \mathrm{E}\left[|x_l(n)|^2 - p\right]^2 + \sum_{l \neq k}^{K} \|p - p_l^{loc}\|^2. \tag{3.2.6}$$

Node k may not have access to all the data p_l^{loc} in the network. We modify the second member of right hand side (RHS) of (3.2.6) by replacing the summation $\sum_{l \neq k}^{K}$ with $\sum_{l \in \mathcal{N}_k/\{k\}}$. Next we replace $\|p - p_l^{loc}\|^2 \approx b_{l,k}\|p - p_l^{loc}\|^2$ ([Say12, Eq. 117]). We collect the non-negative coefficients $b_{l,k}$ in a $K \times K$ matrix $\mathbf{B}$ and assume $b_{l,k} = 0$ if $l \neq \mathcal{N}_k$. Also we replace the unknown p_l^{loc} with an intermediate estimate $\hat{\psi}_l$ available at node l. Then the approximation of (3.2.6) at node k is given as

$$J_k^{dist}(p) = \sum_{l \in N_k} c_{l,k} \, \mathrm{E}\left[|x_l(n)|^2 - p\right]^2 + \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k}\|p - \hat{\psi}_l\|^2 \tag{3.2.7}$$

and derivative of the cost function is (3.2.7) is

$$\nabla_p J_k^{dist}(p) = 2 \sum_{l \in N_k} c_{l,k} \left[p - \mathrm{E}\left[|x_l(n)|^2\right]\right]$$

$$+ 2 \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k} \left[p - \hat{\psi}_l \right]. \tag{3.2.8}$$

The cost (4.3.18) can be used to obtain a recursion for the estimate of p at node k, denoted as $\hat{p}_k(n)$. Using the steepest descent method, which is divided into two parts, we get an iterative solution for (3.2.7) as follows:

$$\begin{aligned} \hat{\psi}_k(n+1) &= \hat{p}_k(n) + \mu_k \sum_{l \in N_k} c_{l,k} \left[\mathrm{E}\left[|x_l(n)|^2 \right] - \hat{p}_k(n) \right] \\ \hat{p}_k(n+1) &= \hat{\psi}_k(n+1) + \nu_k \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k} \left[\psi_l - \hat{p}_k(n) \right]. \end{aligned} \tag{3.2.9}$$

Different step sizes μ_k and ν_k at the nodes k have been assigned and the constants 2 has been incorporated into μ_k and ν_k. In the second equation of (3.2.9) we replace $\hat{\psi}_l$ with time dependant $\hat{\psi}_l(n+1)$, $\hat{p}_k(n)$ with $\hat{\psi}_k(n+1)$ and we get

$$\begin{aligned} \hat{p}_k(n+1) = &\left[1 - \nu_k \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k} \right] \hat{\psi}_k(n+1) \\ &+ \nu_k \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k} \hat{\psi}_l(n+1). \end{aligned} \tag{3.2.10}$$

Next we introduce the coefficients $a_{l,k} = 0$ if $l \neq \mathcal{N}_k$, $a_{l,k} = \nu_k b_{l,k}$ if $l \neq k$ and $a_{k,k} = 1 - \nu_k \sum_{l \in \mathcal{N}_k/\{k\}} b_{l,k}$ if $l = k$. If we collect the coefficients $a_{l,k}$ into a $K \times K$ matrix $\mathbf{A}$, it is straightforward to see that $\sum_{l \in \mathcal{N}_k} a_{l,k} = 1$ for every $k \in K$ and thus $\mathbf{A}$ is a left stochastic matrix [1] (but $\mathbf{A}$ can be also doubly stochastic). We replace $\mathrm{E}\,|x_l(n)|^2$ with $|x_l(n)|^2$ and finally arrive to the Adapt and Combine (ATC) recursions that we summarise with energy detection as Algorithm 1.

In the ATC diffusion algorithm, during the incremental step, at time instant n, the estimate $\hat{\psi}_k(n+1)$ at node k is calculated using the estimate $\hat{p}_k(n)$ at node k and the new observation available for node k. The coefficients $c_{l,k}$ define how the measurements are exchanged between the nodes. During the diffusion step the estimate $\hat{p}_k(n+1)$ at every node k is calculated using a linear combination of the estimates $\hat{\psi}_l(n+1)$ available for node k. The elements $a_{l,k}$ specify the combination strategy of estimates.

Note that in practice the non-negative coefficients $a_{l,k}$ and $c_{l,k}$ can be chosen freely under the conditions, that $\mathbf{C1} = \mathbf{1}$, $\mathbf{1}^T\mathbf{C} = \mathbf{1}^T$, $\mathbf{1}^T\mathbf{A} = \mathbf{1}^T$, $a_{l,k} = 0$, if $l \neq \mathcal{N}_k$ and $c_{l,k} = 0$ if $l \neq \mathcal{N}_k$. The coefficients $b_{l,k}$ are absorbed into coefficients $a_{l,k}$ and do not have to be considered in practice.

[1] For a left stochastic matric $\mathbf{A}$ it holds $\mathbf{1}^T\mathbf{A} = \mathbf{1}^T$.

Algorithm 3.1: Distributed ATC Diffusion Power Estimation

Start with $\hat{p}_k(0) = p(0)$.
Given non-negative real coefficients $a_{l,k}, c_{l,k}$
for every time instant $n \geq 1$ **do**
  for every node $k = 1, ..., K$ **do**
    1. Power estimation:
    $\hat{\psi}_k(n+1) = \hat{p}_k(n)$
      $+\mu_k \sum_{l \in N_k} c_{l,k} \left(|x_l(n)|^2 - \hat{p}_k(n)\right)$
    $\hat{p}_k(n+1) = \sum_{l \in N_k} a_{l,k} \hat{\psi}_l(n+1)$.
    2. Detection decision:
    $H_0 : \hat{p}_k(n+1) < \gamma$ or $H_1 : \hat{p}_k(n+1) > \gamma$.
    (Refer to (3.3.41) for selecting the threshold).
  end for
end for

We also add that if we replace the order of adaptation and fusion equations in (3.2.9) as follows

$$\hat{\psi}_k(n) = \hat{p}_k(n) + \nu_k \sum_{l \in \mathcal{N}_k/\{k\}} b_{k,l} \left[\psi_l - \hat{p}_k(n)\right]$$
$$\hat{p}_k(n+1) = \hat{\psi}_k(n) + \mu_k \sum_{l \in N_k} c_{k,l} \left[\mathrm{E}\,|x_l(n)|^2 - \hat{p}_k(n)\right]. \tag{3.2.11}$$

then by skiping the standard steps, we arrive to the CTA (Combine and Adapt) version of the diffusion LMS algorithm, which is summarised at next In the CTA

Algorithm 3.2: Distributed CTA Diffusion Power Estimation

Start with $\hat{p}_k(0) = p(0)$.
Given non-negative real coefficients $a_{k,l}, c_{k,l}$
for every time instant $n \geq 1$ **do**
  for every node $k = 1, ..., K$ **do**
    1. Power estimation:
    $\hat{\psi}_k(n) = \sum_{l \in N_k} a_{k,l} \hat{p}_l(n)$
    $\hat{p}_k(n+1) = \hat{\psi}_k(n)$
      $+\mu_k \sum_{l \in N_k} c_{k,l} \left(|x_l(n)|^2 - \psi_k(n)\right)$.
    2. Detection decision:
    $H_0 : \hat{p}_k(n+1) < \gamma$ or $H_1 : \hat{p}_k(n+1) > \gamma$.
    (Refer to (3.3.41) for selecting the threshold).
  end for
end for

diffusion algoritm, the estimates $\{\hat{p}_k(n)\}_{k \in \mathcal{N}_k}$ including the $\hat{p}_k(n)$ from node k are combined together at every node k. This is the diffusion step. Then the combined estimate $\hat{\psi}_k(n)$ at node k is used to calculate the new estimate $\hat{p}_k(n+1)$ at node k, using the new observation available for node k, at time instant n. This is the incremental step.

3.2.3 Recursive ring-around topology

As shown in paper [ATB14c], a recusive estimator can be interpreted also as a counterpart of a non-recursive sample variance estimator. By taking into account the suggestions in [Kay93] for a local, non-cooperating estimator for sample variance, the distributed estimator using a circular estimation topology can be constructed as follows

$$\hat{p}_k(n) = \frac{1}{n}\sum_{i=1}^{n} |x_{(k-i+1)\mathrm{mod}K}(n-i+1)|^2. \tag{3.2.12}$$

A recursive equivalent to (3.2.12) is given by

$$\begin{aligned}\hat{p}_k(n) \quad &= \quad \hat{p}_{(k-1)\mathrm{mod}K}(n-1) + \mu(n)(|x_k(n)|^2 \\ &\quad\quad -\hat{p}_{(k-1)\mathrm{mod}K}(n-1)),\end{aligned} \tag{3.2.13}$$

where $n \geq 1$ and with step size: $\mu(n) = \frac{1}{n}$

The usage of step size $\mu(n) = \frac{1}{n}$ however, expects that the received signal $x_k(n)$ over $n \in N$ stays under a fixed hypotheses: H_0 or H_1. This fact makes its direct use in real-time spectrum sensing problematic. As a solution, a positive constant step size $\mu(n) = \mu$ can be used in recursive power estimation algorithm and then (3.2.13) is able to track the possible changes in power of the received signal $x_k(n)$. As common in the literature of adaptive filtering, the step size of the algorithm is user defined.

The estimated power level $\hat{p}_k(n)$ is used as the test statistic of the recursive ED. i.e. $T(x) = \hat{p}_k(n)$. Since there is no fusion centre and for system redundancy purposes, information overhead is allowed in the network. Thus there are K circular estimation processes running in parallel to provide a global estimate for every node $k \in K$. Every node can then perform the energy detection at any time instant. The algorithm can in principle run infinitely (no window for sample processing is required). The proposed algorithm is summarized in Algorithm 3.3. Let us note, that with the suggested algorithm, only one-directional communication with the adjacent node is required for exchanging the soft information, compared to the schemes, where a central processing unit is used and thus two way communication direction is needed to also send the global soft information back to the nodes at every iteration n. An example with $K = 2$ nodes and thereby 2 estimation processes (red and blue) is illustrated in Fig. 3.1 with nodes $k = 1, 2$ receiving samples $n = 1, ..., 3$.

According to AS3 it is assumed, that the number of iterations performed with the recursive algorithm is larger than the number of nodes in the network. The estimator needs to converge to steady state before the detection decision is made

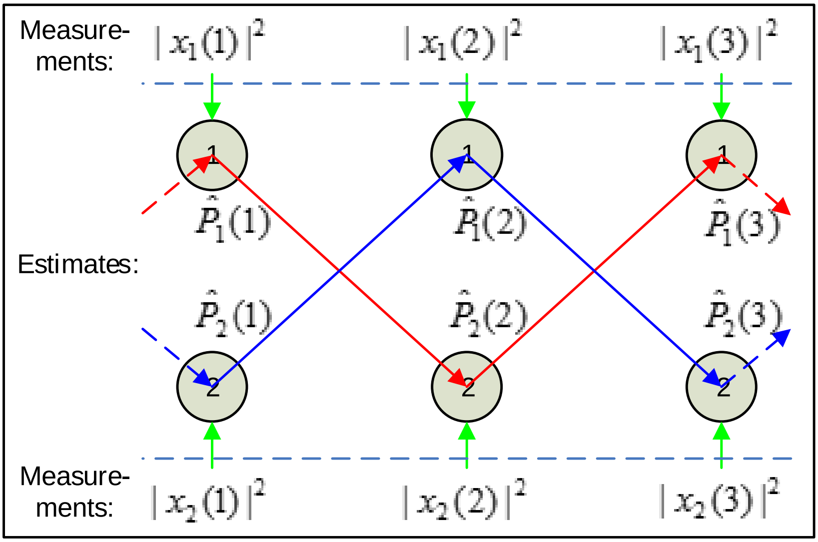

Figure 3.1: Distributed Power Estimation with 2 nodes.

and for the convergence a sufficient number of samples are required. In slow fading the channel coherence time is large and the convergence is achievable. Secondly, in the performance section of the proposed algorithm the Central Limit Theorem (CLT) is applied so enough samples are required also for this approximation to hold. The minimum number of samples for the CLT approximation has been evaluated in the literature, e.g. in [PUP02].

3.2.4 Network topologies

Thus in the ring-around topology [ATB14c], the power estimates are exchanged circularly between the nodes. At time instant n, node k has access only to one estimate $\hat{p}_{(k-1)\text{mod}K}(n)$ from the node $(k-1)\text{mod}K$ for calculating $\hat{p}_k(n+1)$. The local estimate $\hat{p}_k(n)$ is ignored. The algorithm uses only locally observed measurements (i.e $\mathbf{C} = \mathbf{I}$). Thus K estimates have to be sent over the wireless links at time instant n.

Secondly to improve the link failure resistance but keep the need for exchanging the data over wireless links in the network minimal, we compose the diffusion topology from the local $(\mathbf{A}, \mathbf{C} = \mathbf{I})$ and ring-around topologies. At time instant n, at node k the local estimate $\hat{p}_k(n)$ and the estimate $\hat{p}_{(k-1)\text{mod}K}(n)$ from node

Algorithm 3.3: Distributed Ring−Around Power Estimation

Start with $\hat{P}_k(0) = P_0$.
for every time instant $n \geq 1$ **do**
  for every node $k = 1, ..., K$ **do**
    1. Power estimation:
    $\hat{P}_k(n) = \hat{P}_{(k-1)\text{mod}K}(n-1) +$
    $\mu(|x_k(n)|^2 - \hat{P}_{(k-1)\text{mod}K}(n-1))$.
    2. Detection decision:
    $H_0 : \hat{P}_k(n) < \gamma$ or $H_1 : \hat{P}_k(n) > \gamma$.
    (Refer to (3.3.41) for selecting the threshold).
  end for
end for

$(k-1)\text{mod}K$ are fused together using equal, constant weight 0.5 for calculating $\hat{p}_k(n+1)$. For example when $K = 3$ and keeping the same notation and conditions for the elements of matrix $\mathbf{A}$, the ring around and diffusion topologies are given as follows

$$\mathbf{A}^T_{\text{ring}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}^T_{\text{diff}} = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}. \tag{3.2.14}$$

and is illustrated in Fig. 3.2 If measurements are exchanged between the nodes, then we set $\mathbf{C} = \mathbf{A}^T_{\text{diff}}$. Hence at time instant n additionally K measurements have to be exchanged in the network. Otherwise $\mathbf{C} = \mathbf{I}$. Therefore in the subsequent sections we assume, that both matrices $\mathbf{C}$ and $\mathbf{A}$ are doubly stochastic (i.e we have additionally $\mathbf{A}\mathbf{1} = \mathbf{1}$) and all the conditions for selecting elements $a_{l,k}$ and $c_{l,k}$, listed in last subsection, are satisfied.

3.3 Performance analysis

The performance analysis of the proposed algorithms is divided into two parts. First we derive a general model for analysing the mean and variance of the estimates of the ATC, CTA [ATB14a] and ring-around [ATB14c] algorithms in one framework. Next we analyse the resulting energy detection performance. Let us note that for the theoretical performance analysis we need to know the values of the channel gains.

For more convenient notation we stack the estimates and observations from all the nodes $k \in K$ into $K \times 1$ time dependent vectors $\hat{\mathbf{p}}(n) = [\mathbf{p}_1(n) \dots \mathbf{p}_K(n)]^T$ and $\mathbf{x}(n) = \left[|x_1(n)|^2 \dots |x_K(n)|^2\right]^T$ respectively.

Let us define additional matrix $\boldsymbol{\mathcal{M}} = \text{diag}\{\mu_1, \dots, \mu_K\}$, which contain the algorithm step size parameters. We introduce also two additional $K \times K$ matrices

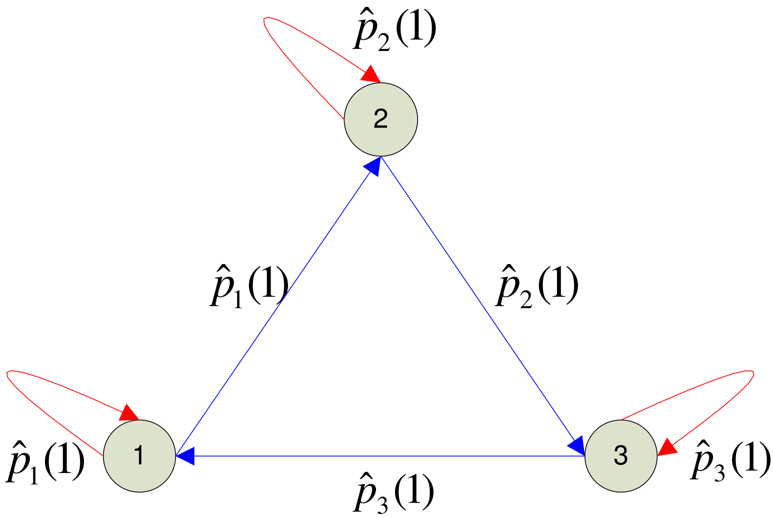

Figure 3.2: Distributed Power Estimation with 3 nodes.

$\mathbf{L}_1$ and $\mathbf{L}_2$ for being able to represent all the 3 algorithms using one framework. Then we can write the recursion in the following general form

$$\hat{\mathbf{p}}(n+1) = \mathbf{L}_2 (\mathbf{I} - \boldsymbol{\mathcal{M}}) \mathbf{L}_1 \hat{\mathbf{p}}(n) + \mathbf{L}_2 \boldsymbol{\mathcal{M}} \mathbf{C} \mathbf{x}(n). \tag{3.3.1}$$

The initial estimate is $\hat{\mathbf{p}}(0)$. It follows, that we get the ATC algorithm, when we take $\mathbf{L}_2 = \mathbf{A}_{\text{diff}}^T$, $\mathbf{L}_1 = \mathbf{I}$, $\mathbf{C} = \mathbf{I}$ or $\mathbf{C} = \mathbf{A}_{\text{diff}}^T$ in case of the measurements are exchanged between the nodes. For CTA algorithm we take $\mathbf{L}_1 = \mathbf{A}_{\text{diff}}^T$, $\mathbf{L}_2 = \mathbf{I}$, $\mathbf{C} = \mathbf{I}$ or $\mathbf{C} = \mathbf{A}_{\text{diff}}^T$. The ring around topology is selected when $\mathbf{L}_2 = \mathbf{I}$, $\mathbf{L}_1 = \mathbf{A}_{\text{ring}}^T$ and $\mathbf{C} = \mathbf{I}$. Note that to keep the matching notation with Algorithm 1, we use transposed matrices in the general recursion. The local, non-cooperative received power estimation is represented by $\mathbf{L}_1 = \mathbf{L}_2 = \mathbf{C} = \mathbf{I}$.

For evaluating the performance of the estimation algorithms and the resulting energy detection, we first evaluate the mean and variance of estimates $\mathbf{p}_k(n)$.

3.3.1 Mean of estimates

Following the signal model (3.2.1), let us denote the conditional expectation of the observation vector as $\mathrm{E}[\mathbf{x}(n)|H_i]$, where $i = 1$ denotes the case when PU signal is present and $i = 0$ the case when PU signal is absent. In this section we assume that the environment is stationary. The conditional means are thus constant over time.

Considering the general recursion (4.4.1), we have

$$\begin{aligned} \mathrm{E}\left[\hat{\mathbf{p}}(n+1)|H_i\right] = \mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\,\mathrm{E}\left[\hat{\mathbf{p}}(n)|H_i\right] \\ +\mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C}E\left[\mathbf{x}(n)|H_i\right], \end{aligned} \tag{3.3.2}$$

for $i = 0, 1$, where the initial value is given as $\mathrm{E}\left[\hat{\mathbf{p}}(0)|H_i\right]$. Let us note that the conditional mean of $\hat{p}_k(n)$ under hypothesis H_i, $i = 0, 1$ at node k is

$$\mathrm{E}\left[\hat{p}_k(n)|H_i\right] = w_k^T\,\mathrm{E}\left[\mathbf{x}(n)|H_i\right] \quad \textit{for i=0,1}, \tag{3.3.3}$$

where $w_k = \mathrm{col}(0\ldots, [w_k(k) = 1], \ldots 0)$ at node k.

After iterating we see, that the mean recursion can be given in the following equivalent form

$$\begin{aligned} \mathrm{E}\left[\mathbf{x}(n)|H_i\right] = \left[\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\right]^n\hat{\mathbf{p}}(0) \\ +\left[\sum_{i=0}^{n-1}\left[\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\right]^i\right]\mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C}E\left[\mathbf{x}(n)|H_i\right]. \end{aligned} \tag{3.3.4}$$

We are interested in finding the mean of the estimates, when the filter has converged to a steady state, i.e when $n \to \infty$. Thus according to (3.3.4) we need to analyse the asymptotic behaviour of $\left[\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\right]^n$ and the limit of the geometric series $\sum_{i=0}^{n-1}\left[\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\right]^i$.

According to [HJ12, Lemma 5.6.11], if for a matrix norm it holds that

$$\|\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\| < 1 \tag{3.3.5}$$

then $\lim_{n\to\infty}[\mathbf{L}_2(\mathbf{I}-\boldsymbol{\mathcal{M}})\mathbf{L}_1]^n \to 0$. Thus given the doubly stochastic matrices $\mathbf{L}_1$, $\mathbf{L}_2$ and $\mathbf{C}$, the choice of step sizes in $\boldsymbol{\mathcal{M}}$ should guarantee that the stability condition (3.3.5) holds. Using the matrix 2-norm and the submultiplicativity property of a matrix norm, we have that

$$\|\mathbf{L}_2\left(I-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\|_2 \le \|\mathbf{L}_2\|_2\|\left(I-\boldsymbol{\mathcal{M}}\right)\|_2\|\mathbf{L}_1\|_2 < 1. \tag{3.3.6}$$

The spectral norm of a doubly stochastic matrix is 1 [2]. Since the matrix $(\mathbf{I}-\boldsymbol{\mathcal{M}})$ is diagonal, we have that

$$\|\mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\|_2 \le \|\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\|_2 = \max_k|1-\mu_k| < 1. \tag{3.3.7}$$

We conclude that for the (3.3.5) to hold, we must select the μ_k, $k = 1\ldots K$ in $\boldsymbol{\mathcal{M}}$ so that the diagonal matrix $(\mathbf{I}-\boldsymbol{\mathcal{M}})$ is stable. Thus we have the following condition

$$|\lambda_k\left[(\mathbf{I}-\boldsymbol{\mathcal{M}})\right]| = |1-\mu_k| < 1 \quad \textit{for all k=1}\ldots K. \tag{3.3.8}$$

[2] See [HJ12, Problem 8.7.P5]

Since in our model we have only one mode of convergence of the filter [Say08], μ_k should be selected in the range:

$$0 < \mu_k < 2. \tag{3.3.9}$$

The geometric series $S_n = \sum_{i=0}^{n-1} \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]^i$, which is generated by matrix $\left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]$, converges if and only if the condition (3.3.5) holds for all λ_i. The condition (3.3.5) guarantees that the $\left[\mathbf{I} - \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]\right]$ is invertible. Thus we can write the geometric series as follows

$$S_n = \left[\mathbf{I} - \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]\right]^{-1} \left[\mathbf{I} - \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]^n\right]. \tag{3.3.10}$$

Hence according to (3.3.5) as $n \to \infty$ the geometric series converges to

$$S_n = \left[\mathbf{I} - \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]\right]^{-1}. \tag{3.3.11}$$

Thus by noting the mean of $\hat{\mathbf{p}}(n)$ in steady state and under both hypotheses H_i, $i = 0, 1$ as $\mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right]$, we can write

$$\begin{aligned} \mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right] = &\left[\mathbf{I} - \left[\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1\right]\right]^{-1} \\ &\times \mathbf{L}_2 \boldsymbol{\mathcal{M}} \mathbf{C}\, \mathrm{E}\left[\mathbf{x}(n)|H_i\right], \end{aligned} \tag{3.3.12}$$

where the conditional expectations of observations $\mathrm{E}\left[\mathbf{x}(n)|H_i\right]$ follow (3.2.1). Similarly to (3.3.3) we have that the mean of $\hat{p}_k(n)$ in steady state is

$$\mathrm{E}\left[\hat{p}_k(\infty)|H_i\right] = w_k^T\, \mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right] \quad \textit{for i=0,1}. \tag{3.3.13}$$

Mean of Ring-Around estimates

Since the iteration cycle of the ring-around estimation structure can be easily tracked, specific results for the ring-round estimates can be given.

The mean of the global estimation recursion (3.2.13) can be found directly. Dropping the mod$K$ notation, we have

$$\mathrm{E}[\hat{p}_k(n)] \quad = \quad (1-\mu)\mathrm{E}[\hat{p}_{k-1}(n-1)] + \mu \mathrm{E}[|x_k(n)|^2]. \tag{3.3.14}$$

The initial condition is $p_0 = \hat{p}_k(0)$. Due to the circular estimation topology we have that $N = KM + m$, where $M = \lfloor N/K \rfloor$ and where m denotes additional iterations after full cycles. Let $\mathrm{E}[\hat{p}_k(N)|H_1]$ denote the mean when PU signal present and $\mathrm{E}[\hat{p}_k(N)|H_0]$ the mean when only noise is present. By iterating recursion (3.3.14), using the proposed notation and replacing the expectations using model (3.2.1), we can write

$$
\begin{aligned}
&\mathrm{E}[\hat{p}_k(N)|H_1] = \\
&\qquad \mu S[\frac{1-(1-\mu)^{KM}}{1-(1-\mu)^{K}} \left[\sum_{l=0}^{K-1}(1-\mu)^l|\alpha_{k-l}|^2\right] \\
&\qquad + \sigma_v^2\left[1-(1-\mu)^{KM+m}\right] \\
&\qquad + p_0(1-\mu)^{KM+m} \\
&\qquad + \mu S\left[(1-\mu)^{KM}\left[\sum_{i=0}^{m-1}(1-\mu)^i|\alpha_{k-i}|^2\right]\right]. \qquad (3.3.15)
\end{aligned}
$$

In line 2 of (3.3.15), the geometric series $\sum_{i=0}^{M-1}(1-\mu)^{Ki}$ has been replaced with its sum. Let us note, that according to lines 2 and 5 of (3.3.15), the mean differs from node to node due to the values and processing order of $|\alpha_k|^2$. When only noise is present then $S = 0$ and

$$
\begin{aligned}
&\mathrm{E}[\hat{p}_k(N)|H_0] = \\
&\qquad p_0(1-\mu)^{KM+m} + \sigma_v^2\left[1-(1-\mu)^{KM+m}\right]. \qquad (3.3.16)
\end{aligned}
$$

According to AS3, $M >> K$ and in steady state of the estimator, when $M \to \infty$, the exponential factors $(1-\mu)^{KM+m}$ and $(1-\mu)^{KM}$ in (3.3.15) converge to 0 if $0 < \mu < 1$. In steady-state, formula (3.3.15) goes to

$$
\begin{aligned}
&\mathrm{E}[\hat{p}_k(\infty)|H_1] = \\
&\qquad \sigma_v^2 + \frac{\mu S}{1-(1-\mu)^K}\left[\sum_{l=0}^{K-1}(1-\mu)^l|\alpha_{k-l}|^2\right] \qquad (3.3.17)
\end{aligned}
$$

and in the noise only case correspondingly to

$$
\mathrm{E}[\hat{p}_k(\infty)|H_0] = \sigma_v^2 \quad . \qquad (3.3.18)
$$

3.3.2 Variance of estimates

Let us denote the conditional covariance of the estimates under the hypothesis H_i, $i = 0, 1$ as $\mathrm{Cov}\left[\hat{\mathbf{p}}(n+1)|H_i\right]$. Similarly let $\mathrm{Cov}\left[\mathbf{x}(n)|H_i\right]$ denote the conditional covariance of the observations. By using recursions (4.4.1), (4.4.2) and standard definition of covariance, taking expectation and considering the fact that $\hat{\mathbf{p}}(n)$ is independent of the observation vector $\mathbf{x}(n)$, it can be shown that the covariance recursion is

$$
\begin{aligned}
\mathrm{Cov}\left[\hat{\mathbf{p}}(n+1)|H_i\right] = \mathbf{L}_2\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_1\,&\mathrm{Cov}\left[\hat{\mathbf{p}}(n)|H_i\right] \\
&\times \mathbf{L}_1^T\left(\mathbf{I}-\boldsymbol{\mathcal{M}}\right)\mathbf{L}_2^T \\
&+ \mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C}\,\mathrm{Cov}\left[\mathbf{x}(n)|H_i\right]\mathbf{C}^T\boldsymbol{\mathcal{M}}\mathbf{L}_2^T. \qquad (3.3.19)
\end{aligned}
$$

where initial estimate of covariance matrix is noted by $Cov\left[\hat{\mathbf{p}}(0)|H_i\right]$, $i = 0, 1$. The covariance matrix of observations $Cov\left[\mathbf{x}(n)|H_i\right]$ is constant over time n.

Next we derive the structure of $K \times K$ covariance matrix $\mathrm{Cov}\left[\mathbf{x}(n)|H_i\right]$. By considering the model (3.2.1), when PU signal is present the main diagonal elements of matrix $\mathrm{Cov}\left[\mathbf{x}(n)|H_1\right]$ – the variances of observations at node $k \in K$ can be shown to be:

$$\mathrm{Var}\left[|x_k(n)|^2|H_1\right] = \left(|\alpha_k|^2\sigma_s^2 + \sigma_{v,k}^2\right)^2. \tag{3.3.20}$$

Similarly when the PU signal is not present and according to AS 2 the variances of observations at node $k \in K$ are given as

$$\mathrm{Var}\left[|x_k(n)|^2|H_0\right] = \sigma_{v,k}^4. \tag{3.3.21}$$

When the PU signal is present, the off diagonal elements of matrix $\mathrm{Cov}\left[\mathbf{x}(n)|H_1\right]$ - the covariance of observations at nodes k and j if $k, j \in K$ and $i \neq j$ can be shown to be:

$$\mathrm{Cov}\left[|x_k(n)|^2, |x_j(n)|^2|H_1\right] = |\alpha_k|^2|\alpha_j|^2\sigma_s^4. \tag{3.3.22}$$

According to AS 2 the noise realizations $v_k(n)$ and $v_j(n)$ are uncorrelated in time and space for $k, j \in K$ and $i \neq j$. Thus when the PU signal is absent the covariance of observations is

$$\mathrm{Cov}\left[|x_k(n)|^2, |x_j(n)|^2|H_0\right] = 0, \tag{3.3.23}$$

for $k, j \in K$ and $i \neq j$.

The variance of $\hat{p}_k(n)$ at node k, given the hypothesis H_i, $i = 0, 1$, can be found by multiplying the recursion (4.4.11) with vector w_k^T from the left and with vector w_k from the right

$$\begin{aligned}\mathrm{Var}\left[\hat{p}_k(n+1)|H_i\right] = w_k^T &\mathbf{L}_2 \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_1 \,\mathrm{Cov}\left[\hat{\mathbf{p}}(n)|H_i\right] \\ &\times \mathbf{L}_1^T \left(\mathbf{I} - \boldsymbol{\mathcal{M}}\right) \mathbf{L}_2^T w_k \\ &+ w_k^T \mathbf{L}_2 \boldsymbol{\mathcal{M}} \mathbf{C} \,\mathrm{Cov}\left[\mathbf{x}(n)|H_i\right] \\ &\times \mathbf{C}^T \boldsymbol{\mathcal{M}} \mathbf{L}_2^T w_k. \end{aligned} \tag{3.3.24}$$

Note that (4.4.11) is in the form of a discrete time algebraic Riccati equation (DARE), [KSH00, App. E]. The steady state variance $\mathrm{Var}\left[\hat{p}_k(\infty)H_i\right]$, $i = 0, 1$, at node $k \in K$ can be recovered by selecting the $\{k, k\}$ element of the steady state covariance matrix $\mathrm{Cov}\left[\hat{\mathbf{p}}(\infty)|H_i\right]$, which has been found as a solution to the DARE. Since the DARE can be solved using standard methods, we skip the details here. We have finally

$$\mathrm{Var}\left[\hat{p}_k(\infty)|H_i\right] = w_k^T \left[\mathrm{Cov}(\hat{\mathbf{p}}(\infty)|H_i)\right] w_k. \tag{3.3.25}$$

To find the solution to DARE we use first the Kronecker product property

$$\mathrm{vec}(U\Sigma V) = \left(V^T \otimes U\right) \mathrm{vec}\left(\Sigma\right) \tag{3.3.26}$$

to vectorize the covariance recursion (4.4.11). The notation vec($\mathbf{A}$) stacks the columns of its matrix argument $\mathbf{A}$ on top of each other, while $\text{vec}^{-1}(\text{vec}(\mathbf{A}))$ denotes the inverse operation to recover the matrix argument from the vector input. Thus we can write:

$$\begin{aligned}&\text{vec}(\text{Cov}(\hat{\mathbf{p}}(n+1)|H_i))\\ &\qquad= [\mathbf{L}_2\,(\mathbf{I}-\boldsymbol{\mathcal{M}})\,\mathbf{L}_1 \otimes \mathbf{L}_2\,(\mathbf{I}-\boldsymbol{\mathcal{M}})\,\mathbf{L}_1]\\ &\qquad\quad \times \text{vec}(Cov(\hat{\mathbf{p}}(n)|H_i))\\ &\qquad\quad + [\mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C} \otimes\ \mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C}]\,\text{vec}(\text{Cov}(\mathbf{x}(n)|H_i).\end{aligned} \tag{3.3.27}$$

In steady state, when $n \to \infty$, the $\text{Cov}(\hat{\mathbf{p}}(n+1)|H_i)$ and $\text{Cov}(\hat{\mathbf{p}}(n)|H_i)$ have converged to the same value.

The solution for $\text{vec}(\text{Cov}(\hat{\mathbf{p}}(\infty))|H_i)$, $i = 0, 1$, leads to the following result

$$\begin{aligned}&\text{vec}(\text{Cov}(\hat{\mathbf{p}}(\infty)|H_i))\\ &\qquad= [\mathbf{I} - [\mathbf{L}_2\,(\mathbf{I}-\boldsymbol{\mathcal{M}})\,\mathbf{L}_1 \otimes \mathbf{L}_2\,(\mathbf{I}-\boldsymbol{\mathcal{M}})\,\mathbf{L}_1]]^{-1}\\ &\qquad\quad \times [\mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C} \otimes\ \mathbf{L}_2\boldsymbol{\mathcal{M}}\mathbf{C}]\,\text{vec}(\text{Cov}(x(\infty)|H_i)).\end{aligned} \tag{3.3.28}$$

Thus the steady state variance $\text{Var}(\hat{p}_k(\infty)H_i)$, $i = 0, 1$, at node $k \in K$ can be recovered by selecting the $\{k, k\}$ element of covariance matrix $\text{Cov}(\hat{\mathbf{p}}(\infty)|H_i)$, which has been found as a solution to DARE and is given by (3.3.28). We have finally

$$\text{Var}(\hat{p}_k(\infty)|H_i) = w_k^T \left[\text{vec}^{-1}\left(\text{vec}(\text{Cov}(\hat{\mathbf{p}}(\infty y)|H_i))\right)\right] w_k. \tag{3.3.29}$$

Variance of Ring-Around estimates

Similarly, as for the mean of ring-around estimates, specific results for the variance of ring-round estimates can be given. Since $\hat{p}_{k-1}(n)$ and $|x_k(n)|^2$ are uncorrelated and by dropping the mod$K$ notation, we have

$$\begin{aligned}\text{Var}[\hat{p}_k(n)] &= (1-\mu)^2\text{Var}[\hat{p}_{k-1}(n-1)]\\ &\quad + \mu^2\text{Var}[|x_k(n)|^2].\end{aligned} \tag{3.3.30}$$

Replacing the expectations using model (3.2.1)

$$\begin{aligned}&\text{Var}[\hat{p}_k(KM+m+1)] =\\ &\mu^2 \sum_{i=0}^{M-1} (1-\mu)^{2Ki} \sum_{l=0}^{K-1} (1-\mu)^{2l}\text{Var}(|x_{k-l}(n)|^2)\\ &+\mu^2(1-\mu)^{2KM} \sum_{i=0}^{m-1} (1-\mu)^{2i}\text{Var}(|x_{k-i}(n)|^2)\end{aligned} \tag{3.3.31}$$

Since $x_k(n)$ is CSCG, then according to model (3.2.1) the PU signal is present, $\text{Var}(|x_k(n)|^2) = (S|\alpha_k|^2 + \sigma_v^2)^2$. Let $\text{Var}[\hat{p}_k(N)|H_1]$ denote the variance when the

PU signal present and $\mathrm{Var}[\hat{p}_k(N)|H_0]$ the variance when received signal contains only noise. By iterating (3.3.30), replacing the variances and using the proposed notation, we have that

$$\begin{aligned}
&\mathrm{Var}[\hat{p}_k(N)|H_1] = \\
&\qquad \mu\sigma_v^4 \frac{1-(1-\mu)^{2(KM+m)}}{2-\mu} \\
&\qquad + \mu^2 \frac{1-(1-\mu)^{2KM}}{1-(1-\mu)^{2K}} \\
&\qquad\quad \cdot \left[\sum_{l=0}^{K-1} (1-\mu)^{2l} \left[S^2|\alpha_{k-l}|^4 + 2S|\alpha_{k-l}|^2\sigma_v^2\right]\right] \\
&\qquad + \mu^2(1-\mu)^{2KM} \\
&\qquad\quad \cdot \left[\sum_{i=0}^{m-1} (1-\mu)^{2i} \left[S^2|\alpha_{k-i}|^4 + 2S|\alpha_{k-i}|^2\sigma_v^2\right]\right].
\end{aligned} \tag{3.3.32}$$

In line 3 of (3.3.32), the geometric series $\sum_{i=0}^{M-1}(1-\mu)^{2Ki}$ has been replaced with its sum. Similarly to the mean, the variance differs over the nodes. When only noise is present, the resulting variance is given as

$$\mathrm{Var}[\hat{p}_k(N)|H_0] = \frac{\mu\sigma_v^4}{2-\mu}\left[1-(1-\mu)^{2(KM+m)}\right]. \tag{3.3.33}$$

In steady state of the estimator, when $M \to \infty$, the exponential factors $(1-\mu)^{2(KM+m)}$ and $(1-\mu)^{2KM}$ in (3.3.32) converge to 0 if the constant step size μ is taken sufficient. Thus the variance tends to

$$\begin{aligned}
\mathrm{Var}[\hat{p}_k(\infty)|H_1] &= \frac{\mu\sigma_v^4}{2-\mu} \\
&+ \frac{\mu^2}{1-(1-\mu)^{2K}} \left[\sum_{l=0}^{K-1} (1-\mu)^{2l} \left[S^2|\alpha_{k-l}|^4 + 2S|\alpha_{k-l}|^2\sigma_v^2\right]\right]
\end{aligned} \tag{3.3.34}$$

under H_1 and in the noise only case to

$$\mathrm{Var}[\hat{p}_k(\infty)|H_0] = \frac{\mu\sigma_v^4}{2-\mu}. \tag{3.3.35}$$

The residual variance of the fixed step size power estimation algorithm depends on the value of μ. We observe, that smaller μ causes smaller residual variance and thus more precise estimation results. On the other hand it is known from the literature of adaptive filtering, that smaller μ causes slower convergence in the mean.

3.3.3 Detection Performance Analysis

The test statistic of the energy detector at node k at time instant n is estimated using distributed received signal power estimation algorithms. Thus the resulting detection performance is dependent on the performance of the underlying estimation process. For deriving the formulas of probability of detection (P_D) and probability of false alarm (P_{FA}) we need to evaluate the probability density function (PDF) of the test statistic $\hat{p}_k(n+1)$ under both hypotheses H_0 and H_1.

The input signal is CSCG and in case $K = 1$, the test statistic of ED $\hat{p}_k(n+1)$ is local and under both hypothesis a Chi-Square distributed random variable with 2N degrees of freedom. The test statistic $\hat{p}_k(n+1)$ is obtained as a sum of a number of identically distributed variables and hence the CLT can be applied to approximate the Chi square distribution by a Gaussian distribution [PUP02]. According to AS 3 the number of samples is large enough, and the CLT is expected to apply.

The global test statistic $\hat{p}_k(n+1)$ in case of hypothesis H_1, is however estimated over independent, but not identically distributed variables. In such a case the Lyapunov CLT [Bil95] can still be applied over a large number of samples to result in a Gaussian approximation.

Let Q be the complementary distribution function of the standard Gaussian

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt. \tag{3.3.36}$$

The conditional mean $\mathrm{E}(\hat{p}_k(n+1)|H_i)$ and the conditional variance $\mathrm{Var}(\hat{p}_k(n+1)|H_i)$ at node k (for $i = 0, 1$), can be easily obtained from previously derived (4.4.2) and (4.4.11) respectively. The conditional moments in steady state can be obtained similarly from the corresponding steady state results. We provide at next approximate formulas for the resulting energy detection performance. The probability of false alarm P_{FA} of the energy detector under hypothesis H_0 is found by

$$P_{FA}(\gamma, t) = Pr(T(x) > \gamma | H_0) = \int_\gamma^\infty p_x(x|H_0) dx \tag{3.3.37}$$

Substituting the estimation mean and variance under H_0, we get

$$P_{FA} = Q\left(\frac{\gamma - \mathrm{E}(\hat{p}_k(n+1)|H_0)}{\sqrt{\mathrm{Var}(\hat{p}_k(n+1)|H_0)}}\right), \tag{3.3.38}$$

which according to AS 2, holds for every node $k \in K$.

The probability of detection of an energy detector under hypothesis H_1 is correspondingly

$$P_D(\gamma, t) = Pr(T(x) > \gamma | H_1) = \int_{\gamma}^{\infty} p_x(x|H_1)dx. \tag{3.3.39}$$

Let the probability of detection at node k be: $P_{D,k}$. Similarly substituting the mean and variance under H_1, we get

$$P_{D,k} = Q\left(\frac{\gamma - \mathrm{E}(\hat{p}_k(n+1)|H_1)}{\sqrt{\mathrm{Var}(\hat{p}_k(n+1)|H_1)}}\right). \tag{3.3.40}$$

The sensing threshold is found from (3.3.38) by fixing the desired value of P_{FA}. Thus

$$\begin{aligned} \gamma \quad = \quad & \mathrm{E}[\hat{p}_k(n+1)|H_0] \\ & + \quad Q^{-1}(P_{FA})\sqrt{\mathrm{Var}[\hat{p}_k(n+1)|H_0]}. \end{aligned} \tag{3.3.41}$$

Due to the AS 2 [ATB14a] the thresholds for every CR node k are equal.

Calculation of the threshold requires, however, knowledge of the moments of the estimation algorithm in case of hypothesis H_0 and these moments are dependent on the algorithm parameters (especially the step size). In practice the required moments can be calculated in advance using (4.4.2) and (4.4.11), known values of the step size and the noise power and then substituting these results into (3.3.41).

3.4 Simulation results

In the numerical simulation section we investigate the ATC power estimation algorithm and compare the results with the CTA [ATB14a] and ring-around [ATB14c] versions. Secondly we view the resulting energy detection performance. In all these simulations the PU signal $s(n)$ is taken as QPSK with unit power S, under the active hypothesis H_1, the step size is: $\mu = 0.01$.

3.4.1 Local and distributed power estimation

We start with investigation of the estimation algorithms. The channel gains are assumed to be constant, fixed during the simulations and obtained by: $\alpha_k \sim \mathcal{CN}(0, 1)$.

Ring-Around

We first investigate the estimates of (3.2.13) under two modes - *local*: if the nodes are not cooperating to each-other (i.e. every node acts as a stand alone energy estimator/detector) and *global*: if nodes are in cooperation. In the next two examples

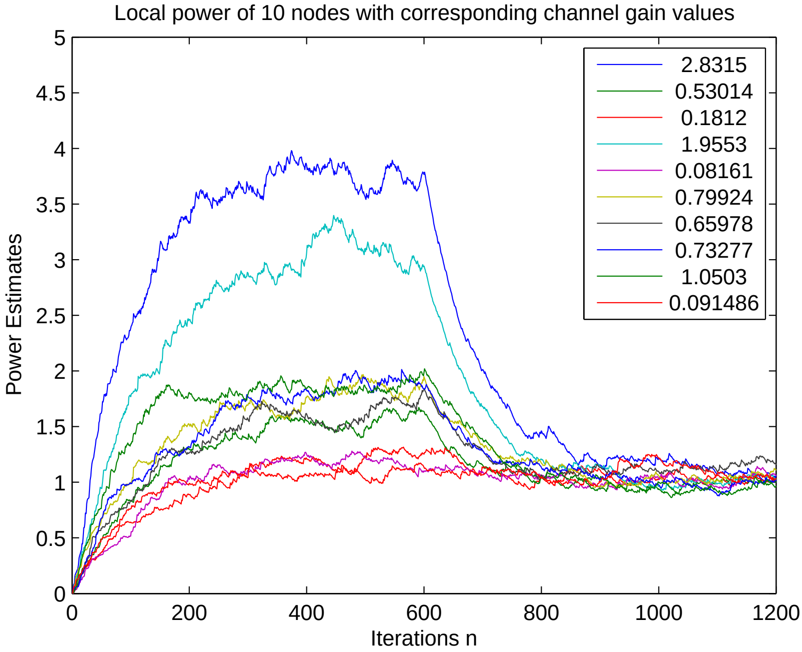

Figure 3.3: Local power estimation, fixed step

all nodes receive $N = 1200$ samples. To illustrate the tracking feature, we examine how the algorithm reacts if the power level changes at sample 601. Thus during samples $n = 1, ..., 600$ hypothesis H_1 is present (the source signal power S is attenuated by channel gain $|\alpha_k|^2$). Due to slow fading the α_k is assumed to be constant and is obtained by: $\alpha_k \sim \mathcal{C}N(0, 1)$. In sample range $n = 601, ..., 1200$, the PU signal is absent and only background noise power $\sigma_v^2 = 1$ is present at every node k.

Using recursion (3.2.13) the local, non-cooperative power estimate is plotted in Fig. 3.3, with 10 nodes in the CR network. The channel gain values $|\alpha_k|^2$ are given on the figure. Obviously, the estimation result using local information is depending on the channel coefficient of the specific node. From $n = 601$ the algorithm is starting to converge to the noise only power level $\sigma_v^2 = 1$. If we for instance chose n dependant step size $\mu = \frac{1}{n}$, then from $n = 601$ the algorithm would obviously not reach to noise level during 600 samples.

In Fig. 3.4 we investigate the cooperative scheme. Exactly the same channel gains are used as in the local simulation. Since the mean and variance differ at nodes k, then for illustration we plot only the global power estimation result of node $k = 10$, in the network with $K = 10$. The corresponding mean and ± 3 times

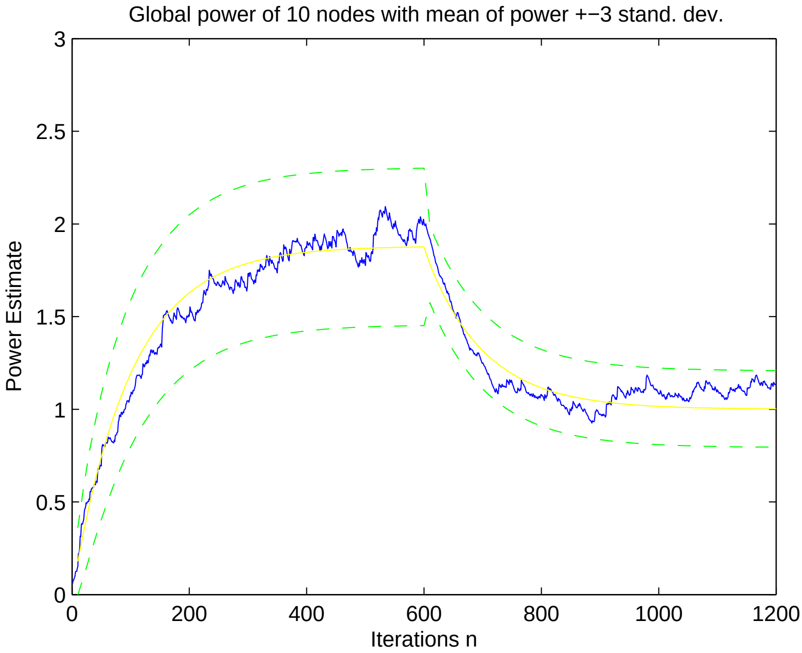

Figure 3.4: Global power estimation, fixed step

standard deviation are given in Fig. 3.4.

In Fig. 3.4 the global estimate is converging around the mean. Due to the proposed circular estimation topology, the recursion (3.2.13) can reduce the effect of random gain caused by channel coefficients. We see that the global estimate stays within the ±3 times standard deviation limits from the mean, which is expected in case of a Gaussian distribution.

ATC and CTA

In the comparison of algorithms we use the same channel gains for all the algorithms. In this subsection, all the nodes in the network receive $N = 2000$ samples. To illustrate how the proposed adaptive algorithms react to changes in the underlying stochastic process, we have changed the active detection hypothesis at sample $n = 1001$. During samples $n = 1 \ldots 1000$ the PU signal with constant unit power S is present. In sample range $n = 1001 \ldots 2000$ the PU signal is absent and only noise is present. Under both detection hypothesis the noise power is $\sigma_v^2 = 1$ and assumed to be the same in all the nodes. In this subsection, it is assumed, that no measurements

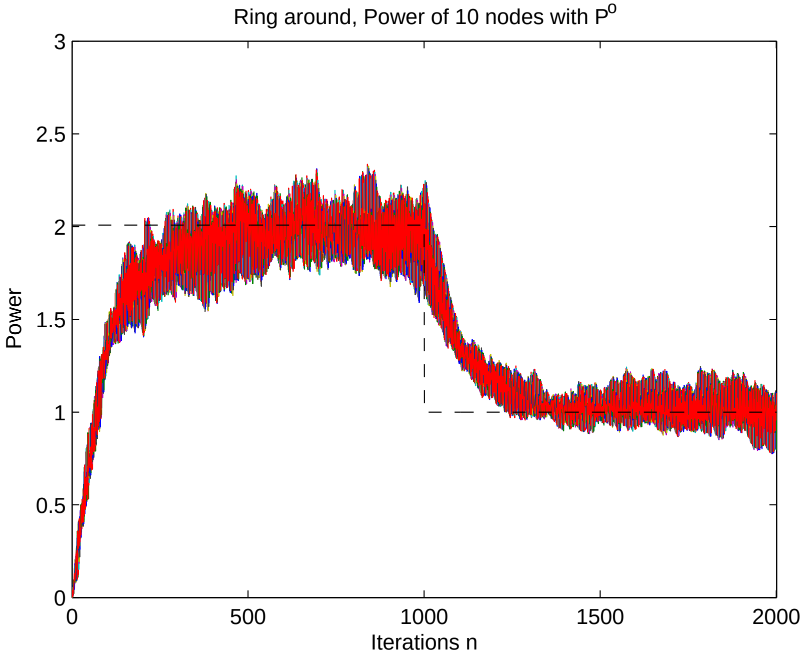

Figure 3.5: Local power estimation.

are exchanged between the nodes, $\mathbf{C} = \mathbf{I}$.

For illustration purpose, all the estimated power values in the CR network of 10 nodes are first plotted in Fig. 3.6. Using the ATC algorithm the estimates of the received power together with the optimal solution P^o have been plotted in Fig. 3.6. All the estimated power values in the CR network of the 10 nodes are plotted in one figure. When we use the CTA algorithm we obtain the results, which are given in Fig. 3.7. The value of optimal solution P^o in figure Fig. 3.6 and in Fig. 3.7 is shown as the black dashed line and is calculated according to (4.3.5) using the present channel gains values.

Compared to the ring round topology in diffusion strategies more information is processed at every node k, since neighbour estimate $(k-1)\text{mod}K$ is fused with the local estimate of node k. It was shown in [ATB14a] that the variance of the estimates of the CTA algorithm is lower than the variance of estimates of the ring around algorithm. Based on Fig. 3.6 and in Fig. 3.7 we observe that the variance of the estimates of the ATC algorithm is even slightly lower than the variance of estimates of the CTA algorithm.

The smallest value of steady state variance is achieved using the ATC algorithm.

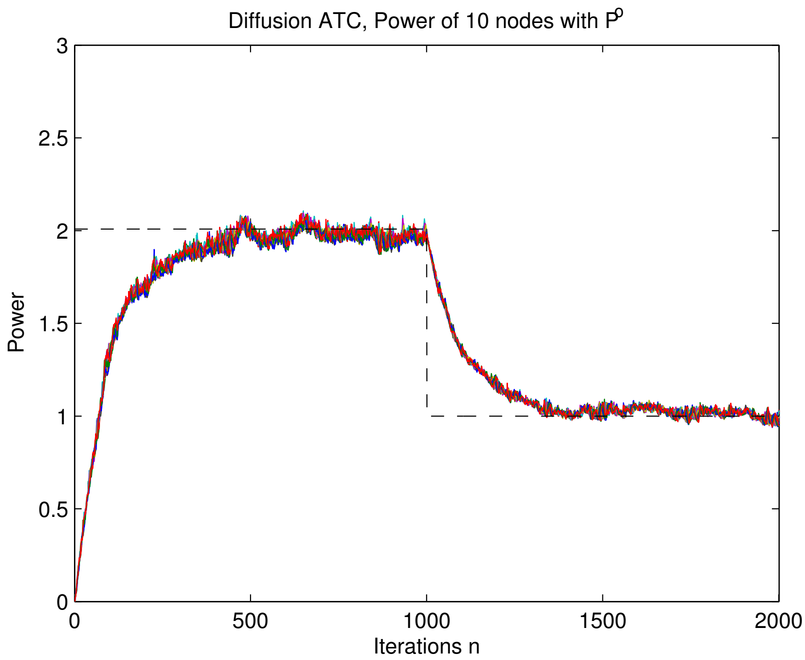

Figure 3.6: Local power estimation using ATC

Compared to the ring around algorithm, since the preciseness of power estimates increases when the diffusion estimation strategies are used, the resulting detection performance will increase as well.

3.4.2 Probability of detection

Next we investigate the probability of detection using the proposed distributed power estimation algorithms. In the following simulations we compare the performance of 5 different network sizes: $K = 1, 3, 10, 30, 50$ nodes. More specifically, the estimated and theoretical results of P_D of the last nodes in the set are compared, i.e $k = K$. In the simulations the converged power estimate is used for detection i.e $\hat{p}_k(\infty)$. The theoretical mean and variance of the power estimates are calculated using directly the steady state formulas.

We set the desired $P_{FA} = 10^{-4}$. The thresholds of the energy detectors at nodes $k \in K$ are calculated using (3.3.41) and the corresponding steady state theoretical mean and variance of the power estimates (of algorithms CTA, ATC and ring around respectively) under detection hypothesis H_0.

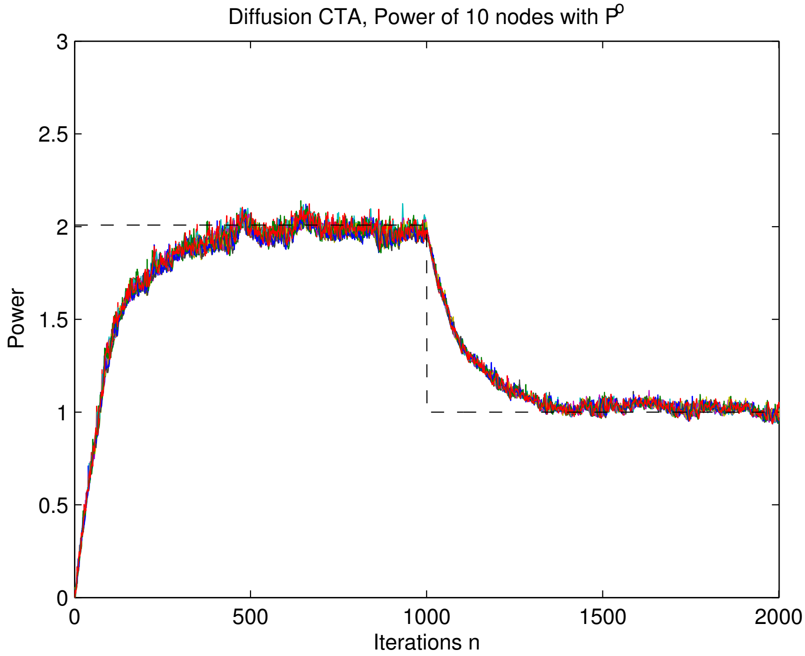

Figure 3.7: Local power estimation using CTA

For estimating the P_D we use the Monte Carlo method [Kay98] and run 1000 experiments with the same fixed set of channel constants and noise power for all the algorithms. The estimated P_D is compared with the theoretical P_D. The theoretical P_D is calculated using (3.3.40) and the corresponding steady state mean and variance of the power estimates of the three algorithms under detection hypothesis H_1. In the following figures the continuous lines represent the theoretical P_D and the corresponding signs the estimated P_D. First we set $\mathbf{C} = \mathbf{I}$. The detection performance of ATC, CTA and the ring around algorithms are shown in Fig. 3.8, in Fig. 3.10 and in Fig. 3.9 respectively. We see that there is a good match between estimated and theoretical P_D. The PDF of the test statistic is approximated by a Gaussian distribution and the CLT approximation applies even with small K and when the underlying stochastic process is cyclostationary (since the variance of the sample of received signal is changing periodically over n). As we noticed in [ATB14a] the CTA algorithm outperforms the ring around algorithm. The P_D of the set with few nodes is more influenced by the given values of channel constants. According to simulation data when $K = 1$ the PU signal is in deep fading and this explains the worse P_D result. In case of non-distributed estimation and detection,

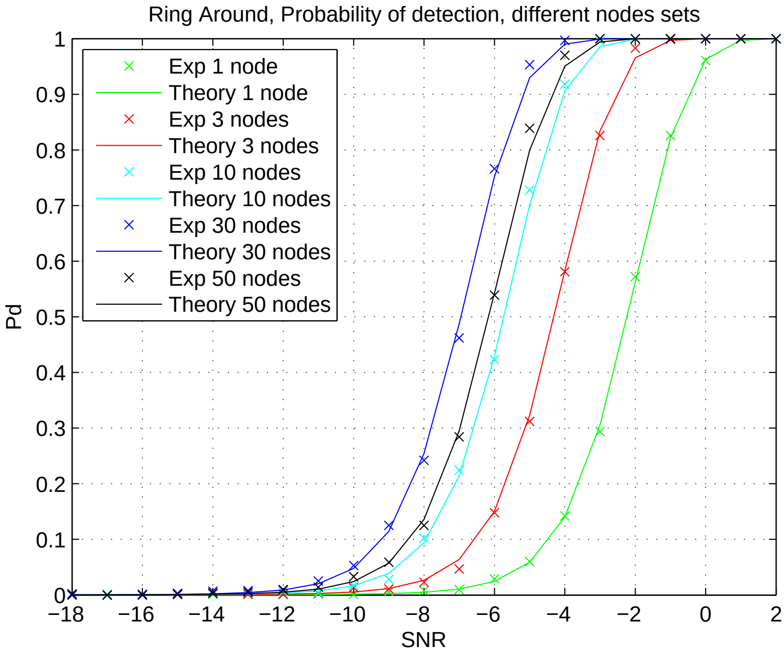

Figure 3.8: Probability of detection, ring around, $\mathbf{C} = \mathbf{I}$

not much can be done to improve the P_D. As the number of nodes in the network increases, about 4 dB is gained with respect to the noise power. Based on Fig. 3.9 we see that the ATC slightly outperforms the CTA. As the number of nodes K increases, from about $K = 30$, the P_D result stabilizes close to the theoretical P_D plot of the no fading case.

When also measurements from a neighbour node are available and we set $\mathbf{C} = \mathbf{A}_{\text{diff}}^T$ for the CTA and ATC algorithms, then the results are shown in Fig. 3.12 and in Fig. 3.11 respectively. We note that ATC performs slightly better, when more nodes in the network. While ATC fuses more data than CTA, the difference of detection performance with CTA is rather small. We see minor increase in the detection performance when additionally measurements are exchanged between the nodes. Thus we conclude that the best detection results are obtained using ATC algorithm, however the difference between ATC and CTA is quite small. On the other hand since exchanging measurements between the nodes in a neighbourhood of a node in the CR network, additional data has to be broadcast, processed and this requires additional energy. Thus the usage of measurement exchange may not be justified in practical implementation.

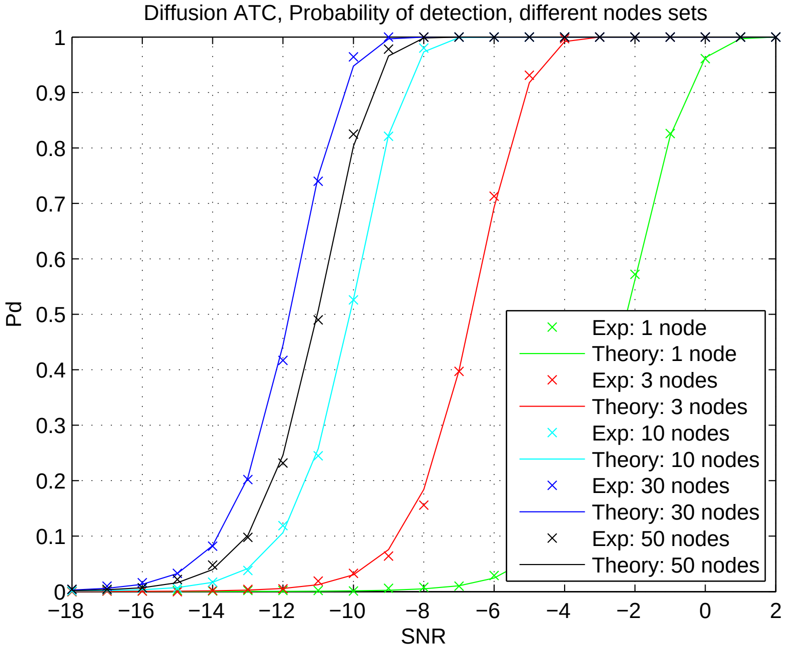

Figure 3.9: Probability of detection, ATC, $\mathbf{C} = \mathbf{I}$

3.5 Conclusion

In this chapter we studied a diffusion based distributed power estimation approach, what is applicable for CR networks for detecting the presence of PU signal. We derived Ring-Around, CTA and ATC diffusion based energy detection algorithms for energy detection. We proposed a general framework for analysing the performance of the ATC diffusion, previously studied CTA and ring-around power estimation algorithms and compared the resulting energy detection performances. Our simulation study demonstrated that both diffusion LMS based energy detection algorithms outperform the previously proposed ring around algorithm and that the ATC diffusion algorithm slightly outperforms the CTA diffusion algorithm and CTA diffusion algorithm outperforms the ring-around algorithm. In addition it was observed that the effect of exchanging measurements in addition to the estimates in CTA and ATC type of algorithms is rather small. All the three proposed algorithms with fixed step size are able to track changes in received signal power and are usable in cognitive radio systems.

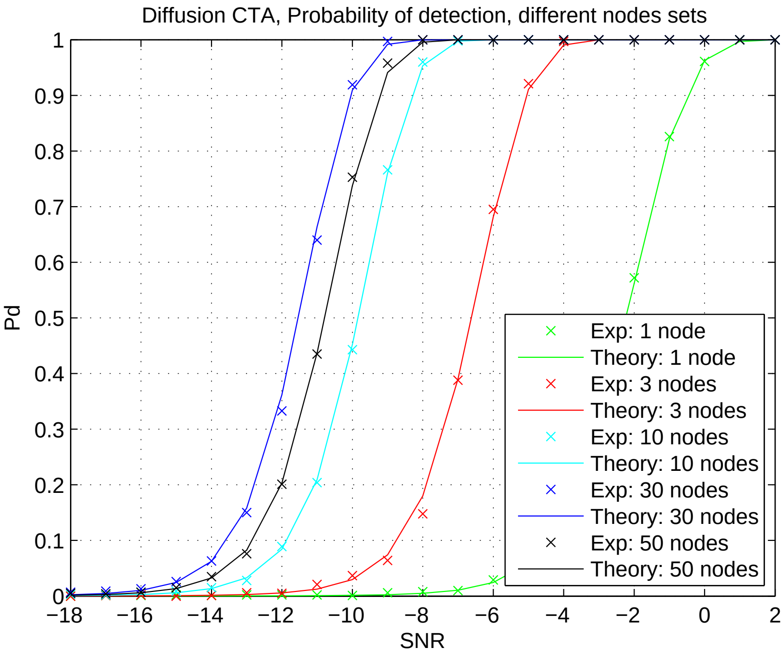

Figure 3.10: Probability of detection, CTA topology, $\mathbf{C} = \mathbf{I}$

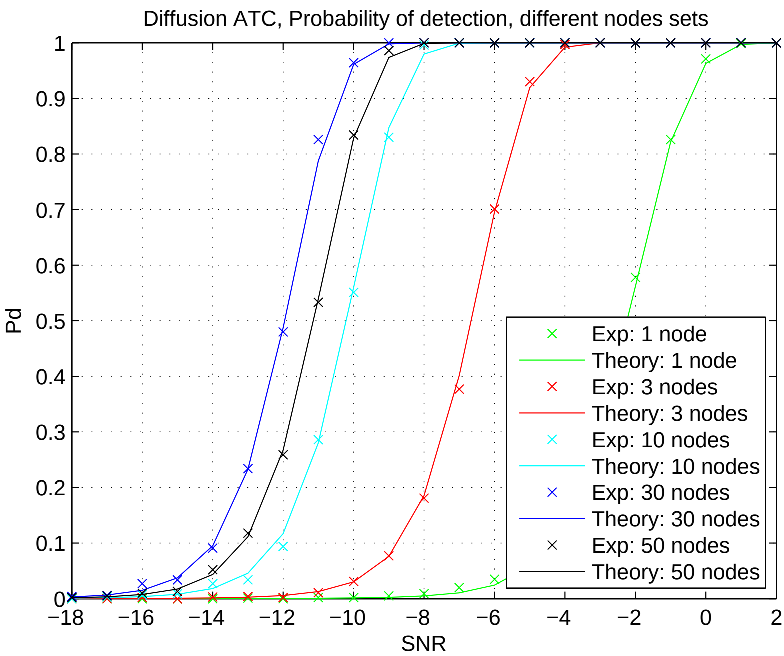

Figure 3.11: Probability of detection, ATC topology, $\mathbf{C} = \mathbf{A}_{\text{diff}}^T$

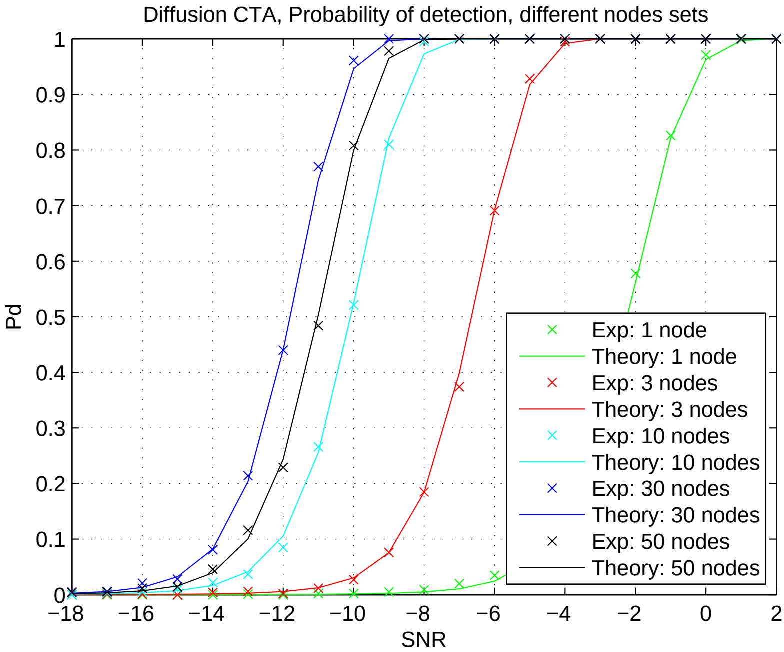

Figure 3.12: Probability of detection, CTA topology, $\mathbf{C} = \mathbf{A}_{\text{diff}}^T$

Chapter 4

Distributed Largest Eigenvalue Based Spectrum Sensing using Diffusion LMS

In this chapter we propose a distributed detection scheme for cognitive radio (CR) networks, based on the largest eigenvalues (LEs) of adaptively estimated correlation matrices (CMs), assuming that the primary user signal is temporally correlated. The proposed algorithm is fully distributed, thereby avoiding the potential single point of failure that a fusion centre (FC) would imply. Different forms of diffusion least mean square (LMS) algorithms are used for estimating and averaging the CMs over the CR network for the LE detection and the resulting estimation performance is analyzed using a common framework. In order to obtain analytic results on the detection performance, the exact distribution of the CM estimates are approximated by a Wishart distribution, by matching the moments. The theoretical findings are verified through simulations.

4.1 Background

We consider the interweave CR paradigm [GSMS09], where CR systems detect the presence of a primary user (PU) signal by sensing the spectrum area of interest. The binary detection problem is studied: PU signal is present or absent [HNZ09, BGG$^+$13, ALLP12]. In the interweave paradigm it is expected that the CR system should accurately detect the transmission of a PU system, when the latter is operating.

As already described in Chapter 3, in the literature several type of detectors for spectrum sensing have been proposed. When the PU signal waveform, channel and additive noise properties are known *a priori*, then the matched filter detector (MFD) is optimal [Kay98]. The MFD requires perfect synchronization between the PU signal waveform and the received signal. However in practice such required knowledge is often not available, which makes the usage of the MFD detector impractical. The cyclostationary feature detection method [Gar91] requires a priori knowledge about the cyclic frequencies of the PU signals, which often is a too strong assumption for practical implementation - in general it is complicated to implement and it requires that knowledge about the type, modulation and configuration properties of the PU signal is available. The Energy Detection (ED) method [Kay98] models the PU signal as a random process and does not require knowledge about the PU signal, modulation type and channel properties. In such a case, when the received PU signal is white, the ED is optimal. However, setting the detection threshold requires knowledge of the noise power value. It has been shown, that if there is uncertainty in the noise power or if the received PU signal is correlated, the ED performance decreases and it is no more optimal [LZPH08].

A second large group of detectors for spectrum sensing are based on eigenvalue properties of an estimated correlation matrix [TW12, WTL14, ZL09]. Detection based on the largest eigenvalue (LE) of estimated CMs [TW12] is optimal when the observations are zero mean Gaussian distributed, we do not have specific information about the PU signal and the channel gains, and when the PU signal is rank one correlated [TNKG10]. The LE method uses knowledge about the additive noise power to determine the detection threshold. Random Matrix Theory has been used to study the performance of the CM eigenvalue based detectors [CD11]. We note, that when linear estimation of CM is used, more sophisticated detectors: the volume based detector (VD) and the covariance based detector (CAV), which avoid eigenvalue or singular value decomposition, have been studied in [HSQ14, HQXZ15] and [ZC09] respectively. Similarly, when linear estimation of a CM is used, several eigenvalue based detectors are robust in the sense, that the noise power value does not influence the test statistics or threshold of the detectors. For example the Eigenvalue Arithmetic to Geometric Mean (AGM) [HFL+15], the Maximum to Minimum eigenvalue ratio (MME), the Energy to Minimum Eigenvalue ratio (EME) [ZL09], the Eigenvalue Moment ratio (EMR) [HFL+15], and the Hadamard [HXZ15] detectors have been proposed in the literature. A method for blind and optimal combination of observations for the ED has been proposed in [ZLZ08]. For these detectors, the performance analysis is based on the assumption that the sample CM is Wishart distributed with known degrees of freedom (DoF), an assumption that does not hold when exponentially weighted (adaptive) CM estimation is used. Also, the proposed approximate or asymptotic analysis of the theoretical detection performance for EME, MME, CAV detectors tend to be inaccurate in the low SNR regime, as seen in [ZL09, ZC09]. Such potential inaccuracy is not well usable for studying the accuracy of distribution parameter approximations of adaptive CM estimates in a low SNR region.

From the previous chapters we know, that in cognitive radio (CR) contexts we would like to avoid creating interference to the PU user and find free spectrum opportunities as fast as possible. On the other hand the active detection hypothesis may change during the processing time. Distributed, adaptive network learning methods, based on exponential averaging estimation, are able to learn the statistical information based on observations received by the nodes in the network. These methods can react to possible changes in the properties of estimated statistics in real time. Several proposed distributed spectrum sensing solutions make use of a central FC. A FC will however form a single point of failure in the network since a malfunction in this unit affects the performance of the whole distributed solution. We therefore propose a CM estimation solution, where the available CM estimates (and corresponding measurements) are fused in cognitive radio network nodes, to allow all nodes to make detection decisions based on data from the neighboring nodes and without involvement of any central processing unit. Such a solution enhances the network failure resistance.

Also in the Chapter 3 we mentioned that several distributed adaptive estimation schemes have been studied in the past. Consensus based schemes are analyzed for example in [XBL06,SMG09,XBL05,DKM+10]. Diffusion estimation schemes are studied for instance in [ZTS12,GJSS16], while Least mean square (LMS) and recursive least squares (RLS) schemes in [CS11a,CS10,LS08,LCS08]. It has been shown, that distributed diffusion strategies can often perform better (in terms of faster convergence and lower Mean Square Deviation) and be more stable compared to consensus algorithms [TS12, Say14]. Several detection solutions, based on distributed estimation, have been studied for example in [CS11b,LKSS12,YSS13,MBMS16]. A ring network topology for distributed energy detection without a FC has been suggested in [KGC11]. In [ATB14c] we proposed and analyzed a diffusion LMS based recursive calculation of the test statistics with ring topology for the energy detectors in cognitive radio network. Ring networks are however sensitive to communication link failures. Combine and Adapt (CTA) LMS diffusion based calculation of the test statistics for the energy detectors was studied in [ATB14a] and an Adapt and Combine (ATC) based version was investigated further in [ATB14b].

In this chapter we study the performance of LE detection in a distributed CR network, based on adaptively, distributively estimated CMs, using the completely distributed diffusion LMS strategy. We focus on the distributed detection problem and the analysis of dynamics of the diffusion estimation process is beyond the scope of the chapter and this thesis. We make the assumption that the CR network does not have prior information about the waveform of the PU signal and about the channel gains in the secondary nodes. We assume that the received PU signals samples are temporally correlated. Secondly in general we assume the noise power level is known. Noise power estimation procedures and analysis of the sensitivity to estimation errors falls outside the scope of this chapter. To analyze the detection performance and determine the threshold value, we follow the ideas of [Zha12, Kha89, GN05] and approximate the distribution of the exponentially averaged CM estimate by a Wishart distribution by moment matching. The resulting DoF for the approximate Wishart distribution will depend both on the step size, the network topology, and under H_1 detection hypothesis will depend also on the value of the noise variance parameter. We have therefore focused on the LE based detection, since under H_1 the robustness of alternative detectors like EME, MME, CAV in case of adaptively estimated CMs, is lost anyway. We however provide a simulation with the MME detector, which is a robust detector. In the distributed CR network, every node acts as an independent detector in terms of detection decision making based on the available CM estimates. Due to limited information about the PU signal and the communication channel, the theoretical global estimation model is proposed as a network-average CM (while in practice the CR nodes have only access to the subset of data from the neighbor nodes). We consider the control-level analysis of the proposed distributed CM estimation and LE detection algorithm to be out of scope of the chapter.

We organize the remainder of the chapter as follows. In Section 4.2 we describe the motivation, specify the system models which are analysed further in this chapter and we motivate the usage of the LE detector. In Section 4.3 we derive an adaptive, distributed CM estimation algorithm based on diffusion LMS strategy and summarize the versions of it. In Section 4.4 we analyse the performance of the proposed distributed CM estimation algorithm using a common framework for moment based analysis for all the versions of the Diffusion LMS algorithm. We propose the usage of Total and General Variance based approximations for being able to model the distributions of adaptive CM estimates under both detection hypotheses. Using these results the theoretical false alarm and the detection performance of the LE detector are studied. In Section 4.5 we present our simulations results and verify the theoretical findings.

Notation. In this chapter we use the following notations. Boldface uppercase and lowercase letters denote matrices and vectors, respectively. $\mathrm{E}[\cdot]$, $\mathrm{Var}[\cdot]$, $\mathrm{Cov}[\cdot]$ denote expectation, variance (of a scalar) and covariance operators, respectively. $\mathrm{vec}[\cdot]$ and $\mathrm{vec}^{-1}[\cdot]$ denote conversion from matrix to vector and from vector to matrix. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^c$ denote the vector or matrix transpose, the Hermitian transpose and the complex conjugate, respectively. $\otimes$ denotes the Kronecker product.

4.2 Problem formulation and background

4.2.1 Signal model and assumptions

Assume that K single-antenna CR nodes are independently sensing the communication band of a PU. Let the observation bandwidth of the communication band be denoted as B. A collection of samples of the down converted continuous time signal $z_s(t)$ are collected every T_s seconds, with sampling period $\delta_s < T_s$. As a result every node individually obtains a vector

$$\mathbf{y}_k(n) = [z_s(nT_s), z_s(nT_s - \delta_s), \ldots, z_s(nT_s - (M-1)\delta_s)]\,, \tag{4.2.1}$$

which gives the following observation model for both detection hypotheses

$$\begin{aligned} H_0 &: \mathbf{y}_k(n) = \mathbf{v}_k(n), \\ H_1 &: \mathbf{y}_k(n) = \alpha_k \mathbf{s}(n) + \mathbf{v}_k(n), \end{aligned} \tag{4.2.2}$$

where $k = 1, 2, ..., K$ is the node number, M is the length of the observation vector, and $n = 1, 2, ...N$ is the sample discrete time index. The primary signal $\mathbf{s}(n)$, the noise $\mathbf{v}_k(n)$ and channel gains α_k at node k are assumed to be statistically independent. We additionally assume that the PU signal follows

$$\mathbf{s}(n) \sim CN_M\left(\mathbf{m}_s, \mathbf{\Sigma}_s\right). \tag{4.2.3}$$

Due to the one communication channel assumption between a CR and PU, temporally correlation models of CMs are justified by the signal model 4.2.2. In the performance analysis of the LE detection scheme, the following assumption will be used.

AS 1. The additive noise $\mathbf{v}_k(n)$ is independently and identically distributed ($i.i.d$) circularly symmetric complex Gaussian (CSCG) noise with zero mean and covariance $\mathbf{\Sigma}_{v,k} = \sigma^2_{v,k}\mathbf{I}_M$. In the CR network $\mathbf{v}_k(n)$ is uncorrelated in time and space. We assume the noise power is known a *priori* and has the same power level for all nodes in the CR network.

Under H_1 we have the following $M \times M$ CM model

$$\mathbf{R}_k = \mathbf{R}_{s,k} + \mathbf{\Sigma}_{v,k}. \tag{4.2.4}$$

Let us denote the actually occupied bandwidth (within the observation bandwidth B) as b. Thus the ratio between occupation and observation bandwidths is denoted as $\beta = b/B$ [HBS15] and the rank of the PU signal matrix can be then approximated as $\text{rank}(\mathbf{R}_{s,k}) \approx \lceil \beta M \rceil$. We assume $M > 1$, $\beta < 1$ and then $\mathbf{R}_{s,k}$ has in general a low rank (see also [Buc87]), while $\mathbf{\Sigma}_{v,k}$ is a scaled identity matrix. This property can be used for detecting the PU signal.

4.2.2 Largest Eigenvalue detection

In this chapter, we focus on the LE detector, which is known to follow from the General Likelihood Ratio approach, when AS 1 holds, the received observation vectors obey a Multivariate Complex Gaussian distribution with zero mean, and when the PU signal population covariance matrix $\mathbf{R}_{s,k}$ is rank one [TNKG10]. The LE detector requires low computational complexity and the detection performance analysis is easy to conduct. As seen in [TW12] and in Section 4.4, there exist usable theoretical results for the conditional distributions without asymptotic approximations, which predict the true performance well both in low and high SNR. The LE method is optimal for one PU signal. In the case of higher rank PU signals (i.e more than one PU signal in the network), then the LE detector is no longer optimal, but still usable. We note that all these existing results from the literature for the LE detector hold when estimating the CM using a standard non-weighted sample covariance matrix, resulting in a complex Wishart distribution.

For the distributed adaptive estimation scheme considered here, this latter assumption is no longer true, but as will be shown in Sections 4.4 and 4.5, the distribution can still be well approximated by a complex Wishart distribution. The DoF approximations depend on the parameters of the distributed and adaptive CM estimation algorithm step-size and under H_1 also on the preciseness of the noise power value (AS 1). Extending the analysis to other type of detectors can therefore be done using the existing results in the literature, for example from [HSQ14, HQXZ15, HFL$^+$15, HXZ15]. As seen in Section 4.5, a noise power uncertainty under the detection hypothesis H_1 causes an inaccuracy to the approximated DoF$|H_1$ value. This effect causes a potential inaccuracy in the theoretical detection performance formula of a detector, which requires the DoF$|H_1$ value. However since the threshold of a robust detector is not affected by the noise power perturbations, then such a detector can still be used in the framework of this chapter. Thus to keep the focus of the chapter, we have limited our study to the LE detector, where AS 1 is necessary for the threshold calculation and to illustrate the effect of accuracy of the DoF approximations under both detection hypotheses. Since the LE detector is vulnerable to the noise power value uncertainty, then in Section 4.5 we also provide a simulation with the robust MME detector in the proposed distributed and adaptive CM estimation framework.

Thus an estimate $\hat{\mathbf{R}}_k(n)$ of the CM $\mathbf{R}_k$ is assumed to be available for every node $k \in K$ at time index n. Let us define the eigenvalues of $\hat{\mathbf{R}}_k(n)$ in non-increasing order as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. Every node k detects the presence of a PU signal by independently determining the LE of the locally available estimate $\hat{\mathbf{R}}_k(n)$ and performing the following detection test

$$\lambda_1 \left[\hat{\mathbf{R}}_k(n)\right] \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{LE,k}, \tag{4.2.5}$$

using a threshold $\gamma_{LE,k}$, which is given in Section 4.4.3 by (4.4.28) or (4.4.31).

Next we implement the diffusion LMS based method to derive a distributed adaptive CM based LE detector in the CR network, so that the algorithm: A) is able to react to a possible change in the statistics of observations on line (i.e when the detection hypothesis changes during the observation time) and B) estimates the CMs in a cooperative manner with an averaging effect over the CR network. CR nodes can have access only to a subset of neighbor nodes and no FC unit is used in the CR network.

4.3 Adaptive, Distributed CM estimation and LE detection

Obviously one of the most simple cooperation strategies is where all the CR nodes are able to exchange their local data (estimates or observations) with all the other nodes in the CR network, i.e the network global data is available at every node. However in practice it means that all nodes have to be within hearing distance of all the other nodes and significant amount of data needs to be exchanged and processed over the CR network. Secondly transmitting and processing of (global) data consumes energy, which may drain the batteries of the CR nodes. In this chapter we assume to have a more general network topology model, where nodes only share data with a subset of neighbor nodes and thus no global data is available. Thus we assume that the CR nodes use low power transmitters (i.e a low energy communication, to save the batteries) we also would like to save some energy required for local data processing. This means that while every CR node k still needs to transmit its estimate or observation at a time instant n, other nodes use data of pre-selected neighbor nodes and in such a way some energy can be saved by processing (in an adaptive manner) less data at every CR node.

We first describe local CM estimation, when the CR nodes in the network do not cooperate. Then we propose a global (theoretical) cost function for estimating the CM in a cooperative manner. We assume, that the K nodes in the CR network estimate a vector parameter $\mathbf{p}^o$ in a distributed manner, where nodes rely only on the information, that is available to them. The network topology is assumed to be fixed over the sensing time. We consider a linear, fixed combination of neighbor estimates and measurements at every node k and time instant n. The proposed global cost needs to be approximated in a distributed manner, where no FC, as a potential single point of failure in the system, is used. The derivation of the ATC and CTA type CM estimation algorithm diffusion power estimation algorithm follows the ideas in [Say12, CS10, ATB14b].

4.3.1 Local estimation

When CR nodes do not cooperate, then according to (4.2.4) $\mathbf{\Sigma}_{v,k} = \mathrm{E}\left[\mathbf{v}_k(n)\mathbf{v}_k(n)^H\right]$ and $\mathbf{R}_{s,k} = \mathrm{E}\left[|\alpha_k|^2\mathbf{s}(n)\mathbf{s}(n)^H\right]$. The estimate $\hat{\mathbf{R}}_k(N)$ of CM $\mathbf{R}_k$ based on the observations $n = 1, \ldots, N$ can be obtained (independently, non-adaptively) at every node k for example as

$$\hat{\mathbf{R}}_k(N) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_k(n)\mathbf{y}_k(n)^H, \tag{4.3.1}$$

We continue with the notation, suitable for the adaptive processing, i.e the estimate $\hat{\mathbf{R}}_k(n)$, available at node k at time instant n. In the light of the signal model cases in [SCO13], we consider two specific PU signal models under the detection hypothesis H_1, where $\mathbf{s}(n)$ is a constant or a random variable. Under the different detection hypotheses, the $\hat{\mathbf{R}}_k(n)$ therefore follows the following Wishart distributions [TW12, WTL14, GN00]

$$\begin{aligned} &H_0 : \hat{\mathbf{R}}_k(n) \sim CW_M(N, \tfrac{1}{N}\mathbf{\Sigma}_{v,k}),\\ &H_1 : \hat{\mathbf{R}}_k(n) \sim CW_M(N, \tfrac{1}{N}\mathbf{\Sigma}_{v,k}, \tfrac{1}{N}\mathbf{\Omega}_k) \quad \text{if} \quad \mathbf{m}_s \neq 0,\\ &H_1 : \hat{\mathbf{R}}_k(n) \sim CW_M(N, \tfrac{1}{N}\mathbf{\Sigma}'_k) \quad \text{if} \quad \mathbf{m}_s = 0, \end{aligned} \tag{4.3.2}$$

where N is the degree of freedom (DoF) parameter, $\mathbf{\Sigma}'_k = \mathbf{R}_{s,k} + \mathbf{\Sigma}_{v,k}$, by following the notation in [GN00, Th. 3.5.2] $\frac{1}{N}\mathbf{\Omega}_k = \left[\frac{1}{N}\mathbf{\Sigma}_{v,k}\right]^{-1}\left[\frac{1}{N}\mathbf{E}_k\mathbf{E}_k^H\right]$, and where the non-zero column n of $M \times N$ mean matrix $\mathbf{E}_k$ equals $\mathrm{E}\left[\alpha_k\right]\mathbf{m}_s$. The first case corresponds to the Complex Central Wishart (CCW) under detection hypothesis H_0, with population covariance matrix $\frac{1}{N}\mathbf{\Sigma}_{v,k}$. The second case with the non-centrality matrix $\frac{1}{N}\mathbf{\Omega}$ corresponds to the Complex Non-central Wishart distribution (NCW) under H_1. We denote it as Case 1. The third case corresponds to the Complex Central Correlated Wishart (CCCW) under H_1 with population covariance matrix $\frac{1}{N}\mathbf{\Sigma}'_k$. We denote it as Case 2.

According to (4.2.4), every node k has a unique channel gain α_k from the PU source, which is not known *a priori* for the nodes. When the nodes in the CR network estimate $\mathbf{R}_k$ without cooperating with other nodes, then the estimates of $\mathbf{R}_k$ are (locally) influenced by the individual channel gains of the corresponding nodes. The local SNR at node k is given by

$$\mathbf{SNR}_k = \frac{\mathrm{Tr}\left[|\alpha_k|^2\left(\mathbf{R}_{s,k} + \mathbf{m}_s\mathbf{m}_s^H\right)\right]}{\mathrm{Tr}\left[\mathbf{\Sigma}_{v,k}\right]}. \tag{4.3.3}$$

As seen, some CR nodes achieve better detection performance due to higher channel gains (i.e due to better position in the space) than the other. We are interested in a scheme, where all nodes can achieve similar detection performance, despite of their individual channel gains. The method (4.3.1) expects that N samples are available for calculation of the estimate and is not adaptive in its nature, i.e the CR system is unable to react quickly to a possible change of a detection hypothesis during the observation time N. This may increase the possibility of false alarm or a miss-detection of the PU user and thus also an interference to the PU user. As seen in next chapters, we find an adaptive, exponential (non-equal weighed) averaging based method for estimated the CMs, which is able to learn and react to the changes in the statistics of the CM in real time and needs to store only data from previous iteration.

4.3.2 Global estimation

The CR nodes could cooperate via internal communication links to enhance the detection performance (of the PU signal(s)) at every node k. In the distributed CR network we assume:

- **AS 2**. There is a common control channel available for the CR system for transferring the network level control messages. The communication links between the CR nodes are ideal and not capacity restricted.

- **AS 3**. The CR network is strongly connected (however nodes can directly communicate only with a subset of neighbor nodes).

We propose a model where nodes jointly (and in case of either detection hypothesis) estimate the network average CM, which is denoted as $\mathbf{R}^o$ and defined as follows

$$\mathbf{R}^o = \frac{1}{K}\sum_{k=1}^{K}\mathbf{R}_k^o. \tag{4.3.4}$$

For notational convenience, introduce $M^2 \times 1$ $\mathbf{r}^o = \text{vec}(\mathbf{R}^o)$. Thus we can write

$$\mathbf{r}^o = \frac{1}{K}\sum_{k=1}^{K}\text{vec}(\mathbf{R}_k^o) = \frac{1}{K}\sum_{k=1}^{K}\text{E}\left[\text{vec}\left[\mathbf{y}_k(n)\mathbf{y}_k(n)^H\right]\right]. \tag{4.3.5}$$

Let us define the Hermitian rank one observation matrix $\mathbf{D}_{R,k}(n) = \mathbf{y}_k(n)\mathbf{y}_k(n)^H$ (under both hypothesis) at node k at time instant n. Its $M^2 \times 1$ vectorized form is $\mathbf{d}_{R,k}(n) = \text{vec}\left[\mathbf{D}_{R,k}(n)\right]$. We can decompose the $\mathbf{d}_{R,k}(n)$ into the product of a $M^2 \times M^2$ constant (invertible) complex matrix $\mathbf{T}$ and a $M^2 \times 1$ real vector $\mathbf{d}_k(n)$ as $\mathbf{d}_{R,k}(n) = \mathbf{T}\mathbf{d}_k(n)$, to keep the dimension of the estimated vector minimal in the adaptive recursions. For example, when $M = 2$, then

$$\mathbf{T}\mathbf{d}_k(n) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & 1 & i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{R,k}(n)(1,1) \\ \Re[\mathbf{D}_{R,k}(n)(1,2)] \\ \Im[\mathbf{D}_{R,k}(n)(1,2)] \\ \mathbf{D}_{R,k}(n)(2,2) \end{bmatrix}. \tag{4.3.6}$$

We denote the estimate of the real valued $\text{E}\left[\mathbf{d}_k(n)\right]$ as $\hat{\mathbf{p}}_k(n)$. To construct an adaptive distributed estimation algorithm, we first relate the estimates of $\mathbf{R}_k^o$ and $\mathbf{R}^o$ in (4.3.4) with the minimization of the following global (network-wise) cost function

$$\mathbf{p}^o = \underset{\mathbf{p}}{\text{argmin}}\sum_{k=1}^{K} J_k(\mathbf{p}) = \underset{\mathbf{p}}{\text{argmin}}\sum_{k=1}^{K}\text{E}\,\|\mathbf{d}_k(n) - \mathbf{p}\|^2, \tag{4.3.7}$$

where the vector $\mathbf{p} \in R^{M^2}$ represents the real valued parameters of the CM, to be estimated. Thus $\mathbf{p}^o$ represents the optimal (real valued) CM estimate or is the optimal solution for the minimization of the Mean Square Error (MSE) type of global aggregate cost function $J^{glob}(\mathbf{p})$, which is given as

$$\begin{aligned} J^{glob}(\mathbf{p}) &= \sum_{k=1}^{K} J_k(\mathbf{p}) \\ &= \sum_{k=1}^{K}\text{E}\left[\|\mathbf{d}_k(n)\|^2 - \mathbf{d}_k^T(n)\mathbf{p} - \mathbf{p}^T\mathbf{d}_k(n) + \mathbf{p}^T\mathbf{p}\right]. \end{aligned} \tag{4.3.8}$$

Let us note that compared to the models in [STC+13, CS11b, CS10], in (4.3.8) both the observation and estimation variables are vectors. By differentiating $J^{glob}(\mathbf{p})$ in (4.3.8) with respect to $\mathbf{p}$ and setting the result to zero, we get

$$\nabla_{\mathbf{p}} J^{glob}(\mathbf{p}) = -\sum_{k=1}^{K}\text{E}\left[\mathbf{d}_k^T(n)\right] + K\mathbf{p}^T = 0. \tag{4.3.9}$$

It follows that

$$\mathbf{p}^o = \frac{1}{K}\sum_{k=1}^{K} \mathrm{E}\left[\mathbf{d}_k(n)\right]. \tag{4.3.10}$$

The Hessian of the aggregate cost function is

$$\nabla^2_{\mathbf{p}} J^{glob}(\mathbf{p}) = 2\mathbf{I}_M. \tag{4.3.11}$$

Obviously $J^{glob}(\mathbf{p})$ in (4.3.8) is strongly convex [Say14, C.18] with the unique solution $\mathbf{p}^o$. Also, in case of one node in the CR system ($K = 1$) or when the nodes do not cooperate, then the individual cost $J_k(\mathbf{p})$ is minimized at the point $\mathbf{p}^o_k = \mathrm{E}\left[\mathbf{d}_k(n)\right]$. Since $\nabla^2_{\mathbf{p}} J^{loc}_k(\mathbf{p}) = 2\mathbf{I}_M$ and the individual cost $J^{loc}_k(\mathbf{p})$ is strongly convex, thus $\mathbf{p}^o_k$ is unique as well.

Compared to [STC+13, CS11b, CS10], in this chapter the local costs $J_k(\mathbf{p})$ are individually not minimized at the same global point $\mathbf{p}^o$ due to different channel conditions. However the derivation of the diffusion LMS algorithm still follows the procedure as proposed in these papers. The proposed optimal solution (4.3.7) is similar to the Pareto model, which is analysed in [CS13].

Note that

$$\begin{aligned}\mathbf{R}^o_k &= \mathrm{vec}^{-1}\left[\mathbf{T}\mathbf{p}^o_k\right]\\ \mathbf{R}^o &= \mathrm{vec}^{-1}\left[\mathbf{T}\mathbf{p}^o\right].\end{aligned} \tag{4.3.12}$$

We seek an iterative solution to estimate the $\mathbf{p}^o_k$ and $\mathbf{p}^o$ in a manner, which is adaptive in time, and is fully distributed (cooperative). We propose to use diffusion LMS based distributed solution.

4.3.3 Iterative Diffusion solutions

Let $\mathcal{N}_k$ denote the neighborhood group of node $k \in K$, i.e $\mathcal{N}_k$ defines the set of nodes l which can send data unidirectionally the node k. The node k is assumed to be always connected to itself. For deriving the diffusion LMS algorithm, we define and use the standard matrices $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{C}$ similarly to [CS10], with non-negative elements $a_{l,k}$, $b_{l,k}$ and $c_{l,k}$, that describe how data is exchanged and combined in the network.

Let us start by defining the $K \times K$ right stochastic matrix $\mathbf{C}$ with non-negative elements so that

$$c_{l,k} = 0 \quad \text{if} \quad l \notin \mathcal{N}_k, \quad \mathbf{C}\mathbf{1} = \mathbf{1}, \tag{4.3.13}$$

where $c_{l,k} = 1$ if node l is connected to the node k. The global cost (4.3.8) can be divided into the local cost of over the neighborhood of node k and the sum of local costs of other nodes over their corresponding neighborhoods, and can be given in the following form

$$J^{glob}(\mathbf{p}) = J^{loc}_k(\mathbf{p}) + \sum_{l \neq k}^{K} J^{loc}_l(\mathbf{p}). \tag{4.3.14}$$

The local cost at every node k can be expressed as a weighted combination of the costs of the neighbors of every node k. Thus with the help of non-negative coefficients $c_{l,k}$ the local cost can be given as follows

$$J_k^{loc}(\mathbf{p}) = \sum_{l \in \mathcal{N}_k} c_{l,k} J_l(\mathbf{p}) \tag{4.3.15}$$

and is minimized at the location $\mathbf{p}_k^{loc}$. The following relation $J_l^{loc}(\mathbf{p}) \approx J_l^{loc}(\mathbf{p}^{loc}) + \|\mathbf{p} - \mathbf{p}_l^{loc}\|^2$ [CS12] can be used for the second part of right hand side (RHS) of (4.3.14) to relate the variable $\mathbf{p}$ and the $\mathbf{p}_l^{loc}$. Here the $J_k^{loc}(\mathbf{p}_l^{loc})$, can be ignored, since it is independent on the variable $\mathbf{p}$. Thus we have the modified global cost function $J^{glob'}$ as follows

$$J^{glob'}(\mathbf{p}) = J_k^{loc}(\mathbf{p}) + \sum_{l \neq k}^{K} \|\mathbf{p} - \mathbf{p}_l^{loc}\|^2. \tag{4.3.16}$$

Note that it is not assumed, that node k has access to all the $\mathbf{p}_l^{loc}$ in the network. Thus we need to approximate the $J^{glob'}(\mathbf{p})$ locally at every node k and the standard steps follow. We use the non-negative coefficients $b_{l,k}$ to define if $\mathbf{p}_l^{loc}$ is available for the node k. Thus the elements $b_{l,k}$ take the following values

$$\text{if} \quad l \notin \mathcal{N}_k \quad \text{then} \quad b_{l,k} = 0 \quad \text{else} \quad b_{l,k} = 1. \tag{4.3.17}$$

Then, we limit the summation $\sum_{l \neq k}^{K} \|\mathbf{p} - \mathbf{p}_l^{loc}\|^2$ on the RHS of (4.3.16) to the neighbors of node k i.e $\sum_{l \in \mathcal{N}_k / \{k\}} b_{l,k} \|\mathbf{p} - \mathbf{p}_l^{loc}\|^2$. Secondly, we replace the (only theoretically available) $\mathbf{p}_l^{loc}$ with an intermediate estimate $\hat{\boldsymbol{\psi}}_l$, which is available at node l.

After these steps the approximation of (4.3.16) at node k is given as

$$\begin{aligned} J_k^{dist}(\mathbf{p}) = &\sum_{l \in N_k} c_{l,k} \operatorname{E} \|\mathbf{d}_l(n) - \mathbf{p}\|^2 \\ &+ \sum_{l \in \mathcal{N}_k / \{k\}} b_{l,k} \|\mathbf{p} - \hat{\boldsymbol{\psi}}_l\|^2. \end{aligned} \tag{4.3.18}$$

The steepest descent algorithm [Say08] can be used to obtain a recursion for the estimate of $\mathbf{p}^o$ at time instant n, at node k, denoted as $\hat{\mathbf{p}}_k(n)$. By skipping the derivation steps, as in [CS10], the two-step steepest descent recursions are then given as

$$\hat{\boldsymbol{\psi}}_k(n+1) = \hat{\mathbf{p}}_k(n) + \mu_k \sum_{l \in N_k} c_{l,k} \left[\mathbf{d}_l(n) - \hat{\mathbf{p}}_k(n)\right]$$

$$\hat{\mathbf{p}}_k(n+1) = \left[1 - \nu_k \sum_{l \in \mathcal{N}_k / \{k\}} b_{l,k}\right] \hat{\boldsymbol{\psi}}_k(n+1)$$

$$+\nu_k \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}\hat{\boldsymbol{\psi}}_l(n+1), \tag{4.3.19}$$

where μ_k and ν_k are a positive step sizes, $\hat{\boldsymbol{\psi}}_k(n+1)$ is an intermediate estimate at node k at time n.

The coefficients in front of $\hat{\boldsymbol{\psi}}_l(n+1)$, $l = 1,\dots,K$ in the second equation of (4.3.19) can be incorporated into the non-negative coefficients $a_{l,k}$. Let us introduce the $K \times K$ matrix $\mathbf{A}$, whose elements satisfy

$$a_{l,k} = 0 \quad \text{if} \quad l \notin \mathcal{N}_k, \quad \mathbf{1}^T\mathbf{A} = \mathbf{1}^T. \tag{4.3.20}$$

Thus we take $a_{k,k} = 1 - \nu_k \sum_{l\in\mathcal{N}_k/\{k\}} b_{l,k}$ and $a_{l,k} = \nu_k b_{l,k}$ for $l \neq k$. It is straightforward to see that $\sum_{l\in\mathcal{N}_k} a_{l,k} = 1$ for every $k \in K$ and thus $\mathbf{A}$ is a left stochastic matrix. Finally we obtain the Adapt and Combine (ATC) recursions as

$$\hat{\boldsymbol{\psi}}_k(n+1) = \hat{\mathbf{p}}_k(n) + \mu_k \sum_{l\in N_k} c_{l,k}\left(\mathbf{d}_l(n) - \hat{\mathbf{p}}_k(n)\right)$$
$$\hat{\mathbf{p}}_k(n+1) = \sum_{l\in N_k} a_{l,k}\hat{\boldsymbol{\psi}}_l(n+1). \tag{4.3.21}$$

In similar manner the Combine and Adapt (CTA) version can be derived, following the ideas from [CS10]. In the ATC and CTA algorithms the coefficients $c_{l,k}$ and $a_{l,k}$ define respectively how the measurements $\mathbf{d}_l(n)$ and $\hat{\mathbf{p}}_l(n)$ are (unidirectionally) available for the node k. Thus the matrices $\mathbf{A}$ and $\mathbf{C}$ specify the combination strategy of the measurements and the estimates respectively in the CR network.

In Algorithm 4.1 we present the ATC and CTA based CM estimation recursions and the detection step in a common form. For this we define an additional intermediate estimate $\hat{\boldsymbol{\phi}}_k(n)$ and denote the $K \times K$ matrix $\mathbf{A}$ as $\mathbf{A}_1$ or $\mathbf{A}_2$, with the elements $a_{1,l,k}$ and $a_{2,l,k}$ correspondingly. The selection options of the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ and $\mathbf{C}$ based on [CS10] are given in Table 1. In practice the non-negative coefficients $a_{1,l,k}$, $a_{2,l,k}$, $c_{l,k}$ can be chosen freely under the conditions (4.3.13) and (4.3.20) respectively. The coefficients $b_{l,k}$ are absorbed into coefficients $a_{l,k}$ and do not have to be considered in practice. For comparison in Section 4.5, we list also a topology, where every node acts as a FC, denoted as Global FC LMS in Table 1. In such case CR nodes estimate the CM adaptively and independently (without sharing estimates), all the measurements from all the CR nodes are available and equally weighted for every node in the network.

Thus we observe that according to (4.3.12), Table 1 and the CM estimation recursions in Algorithm 4.1, when the nodes in the CR network do not cooperate, then the adaptive estimate $\hat{\mathbf{p}}_k(n)$ at time instant n at node k defines the individual (local) adaptive estimate of $\mathbf{R}_k^o$. When nodes cooperate by following the proposed cost (4.3.7), Table 1 and the CM estimation recursions in Algorithm 4.1, then the adaptive estimate $\hat{\mathbf{p}}_k(n)$ at time instant n at node k defines the adaptive estimate

Algorithm 4.1: Distributed LMS based CM Estimation and Detection

Start with $\hat{\mathbf{p}}_k(0) = \mathbf{p}(0)$ for every k .
Given non-negative real coefficients $a_{1,l,k}, a_{2,l,k}, c_{l,k}$
for every time instant $n \geq 1$ **do**
  for every node $k = 1, ..., K$ **do**
    1. CM estimation recursions:
    $\hat{\boldsymbol{\phi}}_k(n) = \sum_{l=1}^{K} a_{1,l,k} \hat{\mathbf{p}}_l(n))$.
    $\hat{\boldsymbol{\psi}}_k(n+1) = \hat{\boldsymbol{\phi}}_k(n)$
      $+\mu_k \sum_{l=1}^{K} c_{l,k} \left[\mathbf{T}^{-1}\mathbf{d}_{R,l}(n) - \hat{\boldsymbol{\phi}}_k(n)\right]$
    $\hat{\mathbf{p}}_k(n+1) = \sum_{l=1}^{K} a_{2,l,k} \hat{\boldsymbol{\psi}}_l(n)$
    2. LE detection decision:
    $H_0 : \lambda_1 \left[\text{vec}^{-1}\left[\mathbf{T}\hat{\mathbf{p}}_k(n+1)\right]\right] < \gamma_{LE,k}$ or
    $H_1 : \lambda_1 \left[\text{vec}^{-1}\left[\mathbf{T}\hat{\mathbf{p}}_k(n+1)\right]\right] > \gamma_{LE,k}$.
    (Refer to (4.4.28) or (4.4.31) for selecting the $\gamma_{LE,k}$).
  end for
end for

Table 4.1: Choices of Matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ and $\mathbf{C}$ for different LMS algorithms

Algorithm	$\mathbf{A}_1$	$\mathbf{A}_2$	$\mathbf{C}$
No Cooperation LMS	$\mathbf{I}$	$\mathbf{I}$	$\mathbf{I}$
Global FC LMS [CS10]	$\mathbf{I}$	$\mathbf{I}$	$(1/K)\mathbf{1}\mathbf{1}^T$
CTA diffusion LMS [CS10]	$\mathbf{A}$	$\mathbf{I}$	$\mathbf{C}$
ATC diffusion LMS (4.3.21)	$\mathbf{I}$	$\mathbf{A}$	$\mathbf{C}$

of $\mathbf{R}^o$ in (4.3.4), within acceptable mean square error bounds [CS11b, CS10]. Thus after several iterations, the adaptive estimate $\hat{\mathbf{R}}_k(n)$ of $\mathbf{R}^o$ is available (via the transformation (4.3.6) and de-vectorization) for every node in the CR network. Therefore depending on the cooperation model of the nodes, the node k at time instant n can perform independently the LE detection based on the available matrix estimate $\hat{\mathbf{R}}_k(n) = \text{vec}^{-1}\left[\mathbf{T}\hat{\mathbf{p}}_k(n)\right]$.

Regarding the communication cost of Algorithm 4.1, then based on Table 1 it is obvious, that when $\mathbf{A} \neq \mathbf{I}$, then from the transmission point of view still every node $k \in K$ needs to broadcast its $M^2 \times 1$ estimation vector $\hat{\mathbf{p}}_k(n)$ at time instant n to the neighbours of hearing distance of the node k. However from the receiving point of view the number of estimates $\hat{\mathbf{p}}_k(n)$ required for the fusion by every node k is determined by the selection of matrix $\mathbf{A}$. Similarly, every node k obtains at time instant n a $M^2 \times 1$ observation vector $\hat{\mathbf{d}}_k(n)$ and when $\mathbf{C} \neq \mathbf{I}$ broadcasts it at time instant n to the neighbours of hearing distance of the node k. Thus on the receiving side, the exact selection of $\mathbf{C}$ determines the number $\hat{\mathbf{d}}_k(n)$ required by every node k at time instant n for observation fusion. In Section 4.5.1 we comment our selection of $\mathbf{A}$ and $\mathbf{C}$ for the simulations.

Finally we note that in addition to AS 2, obviously the CR system needs some control layer protocol to establish a connection between the nodes. The details of this operation is outside the scope of this chapter. We note, the exact control layer model and implementation of the CR system is out of scope or the paper. In general a protocol needs to be implemented to control the (iteration) time of reliable spectrum sensing and the time of transmission of secondary (CR) system. Thus based on assumption AS 4, the Algorithm 4.1 is started and running on-line, until stopped or , re-initiated by the system.

4.4 Performance analysis

The performance analysis of the proposed algorithm is divided into three parts. First we derive a general model for analyzing the mean and (co-)variance of the adaptive CM estimates of recursions in Algorithm 4.1 in one framework. Secondly we study the statistical properties of the adaptive CM estimates. For studying the LE detection performance of the adaptive CM estimate, the distribution of the adaptive CM estimate is approximated by a CCCW distribution. We propose the usage of the Total and General Variance methods for approximation the DoF and mean matrix parameters for the corresponding CCCW distributions, based on the moments of adaptive CM estimates. Thirdly we provide theoretical results for the LE detector. Let us note that for the theoretical performance analysis of the LE detector, we need to know the values of the channel gains and the noise power.

4.4.1 Moment analysis of adaptive CM estimates

For the analysis of the moments of the spatio-temporal adaptive CM estimates, we propose to use a more general vector/matrix recursion model.

We stack first the $M^2 \times 1$ estimates and observations from all the nodes $k \in K$ into a $KM^2 \times 1$ column vector $\hat{\mathbf{p}}(n)|H_i = [\hat{\mathbf{p}}_1(n)|H_i \dots \hat{\mathbf{p}}_K(n)|H_i]^T$ and $\mathbf{d}(n)|H_i = [\mathbf{d}_1(n)|H_i \dots \mathbf{d}_K(n)|H_i]^T$ respectively, where $i = 1$ denotes the case when the PU signal is present and $i = 0$ the case when the PU signal is absent. The initial estimate is noted as $\hat{\mathbf{p}}(0)|H_i$.

Secondly we define an additional $K \times K$ matrix $\boldsymbol{\mathcal{M}} = \text{diag}\{\mu_1, \dots, \mu_K\}$, which contains the positive step size parameters of the algorithms for every node $k \in K$. The matrix $\boldsymbol{\mathcal{M}}$ is then be extended to another $KM^2 \times KM^2$ matrix as $\overline{\boldsymbol{\mathcal{M}}} = \boldsymbol{\mathcal{M}} \otimes \mathbf{I}_{M^2}$. For the purpose of comparison with the Consensus algorithm [TS12], let the $K \times K$ matrix $\mathbf{A}_0$ specify the fusion strategy of estimates of the consensus algorithm.

The $K \times K$ network topology matrices $\mathbf{A}_0$, $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{C}$ are extended to $KM^2 \times KM^2$ matrices as follows, $\overline{\mathbf{A}}_0 = \mathbf{A}_0^T \otimes \mathbf{I}_{M^2}$, $\overline{\mathbf{A}}_1 = \mathbf{A}_1^T \otimes \mathbf{I}_{M^2}$, $\overline{\mathbf{A}}_2 = \mathbf{A}_2^T \otimes \mathbf{I}_{M^2}$

and $\overline{\mathbf{C}} = \mathbf{C}^T \otimes \mathbf{I}_{M^2}$.

Proposition 1. The distributed LMS algorithms in Table 1 and the consensus algorithm [TS12] can be described by the following spatio-temporal recursion

$$\hat{\mathbf{p}}(n+1)|H_i = \overline{\mathbf{A}}_2 \left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1 \hat{\mathbf{p}}(n)|H_i + \overline{\mathbf{A}}_2 \overline{\boldsymbol{\mathcal{M}}} \overline{\mathbf{C}} \mathbf{d}(n)|H_i. \tag{4.4.1}$$

In case of LMS algorithms $\mathbf{A}_0 = \mathbf{I}_K$ and for example we get the ATC algorithm with no measurement exchange, when we take additionally $\mathbf{A}_1 = \mathbf{C} = \mathbf{I}_K$ and $\mathbf{A}_2 \neq \mathbf{I}_K$, according to the selected network topology. Thus $\overline{\mathbf{A}}_0 = \overline{\mathbf{A}}_1 = \mathbf{I}_K \otimes \mathbf{I}_{M^2}$, $\overline{\mathbf{A}}_2 = \mathbf{A}_2^T \otimes \mathbf{I}_{M^2}$ and $\overline{\mathbf{C}} = \mathbf{I}_K \otimes \mathbf{I}_{M^2}$. For CTA algorithm we take $\overline{\mathbf{A}}_1 = \mathbf{A}_{\text{diff}}^T \otimes \mathbf{I}_{M^2}$, $\overline{\mathbf{A}}_2 = \mathbf{I}_K \otimes \mathbf{I}_{M^2}$, $\overline{\mathbf{C}} = \mathbf{I}_K \otimes \mathbf{I}_{M^2}$ or $\overline{\mathbf{C}} = \mathbf{A}_{\text{diff}}^T \otimes \mathbf{I}_{M^2}$. Note that to keep the matching notation with Algorithm 1, we use transposed matrices in the general spatio-temporal vector recursion. For the Consensus algorithm [TS12], we take $\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{C} = \mathbf{I}_K$, $\mathbf{A}_0 \neq \mathbf{I}_K$ according to the network topology and thus we have $\overline{\mathbf{A}}_0 = \mathbf{A}_0^T \otimes \mathbf{I}_{M^2}$ and $\overline{\mathbf{A}}_1 = \overline{\mathbf{A}}_2 = \mathbf{I}_K \otimes \mathbf{I}_{M^2}$. Note, that the proposed Kronecker extension retains the stochastic property of the extended matrix and due to the transpose, the matrices $\overline{\mathbf{A}}_1$ and $\overline{\mathbf{A}}_2$ are now right stochastic and $\overline{\mathbf{C}}$ is left stochastic.

For studying the performance of the LMS algorithms, we first need to evaluate the moments - mean and covariance of the stacked estimates $\hat{\mathbf{p}}(n)$ and we provide the corresponding recursions for evaluating these moments.

Mean of estimates

Let us denote the conditional expectation of the observation vector as $\mathrm{E}\left[\mathbf{d}(n)|H_i\right]$, where $i = 0, 1$. We specify these values in the Section 4.4.2.

Proposition 2. The general recursion (4.4.1), can be expressed as

$$\begin{aligned}\mathrm{E}\left[\hat{\mathbf{p}}(n+1)|H_i\right] = \overline{\mathbf{A}}_2 \left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1 \mathrm{E}\left[\hat{\mathbf{p}}(n)|H_i\right] \\ + \overline{\mathbf{A}}_2 \overline{\boldsymbol{\mathcal{M}}} \overline{\mathbf{C}} E\left[\mathbf{d}(n)|H_i\right],\end{aligned} \tag{4.4.2}$$

for $i = 0, 1$, where the initial value for the mean vector is given as $\mathrm{E}\left[\hat{\mathbf{p}}(0)|H_i\right]$, $i = 0, 1$.

After iterating we see, that the mean recursion can be given in the following equivalent form

$$\begin{aligned}\mathrm{E}\left[\hat{\mathbf{p}}(n)|H_i\right] = \left[\overline{\mathbf{A}}_2 \left(\overline{\mathbf{A}}_0 - \overline{\mathcal{M}}\right) \overline{A\mathbf{1}}\right]^n \hat{\mathbf{p}}(0) \\ + \left[\sum_{i=0}^{n-1} \left[\overline{\mathbf{A}}_2 \left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]^i\right] \\ \times \overline{\mathbf{A}}_2 \overline{\boldsymbol{\mathcal{M}}} \overline{\mathbf{C}} E\left[\mathbf{d}(n)|H_i\right].\end{aligned} \tag{4.4.3}$$

For the asymptotic analysis of the mean recursion (4.4.3), we need to analyse the asymptotic behavior of $\left[\overline{\mathbf{A}}_2\left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1\right]^n$ and the limit of the geometric series $\sum_{i=0}^{n-1}\left[\overline{\mathbf{A}}_2\left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1\right]^i$, when $n \to \infty$.

According to [HJ12, Theorem 5.6.12], the convergence $\lim_{n\to\infty}[\overline{\mathbf{A}}_2\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1]^n \to 0$ happens if and only if the spectral radius of the matrix $\overline{\mathbf{A}}_2\left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1$ satisfies

$$\rho\left(\overline{\mathbf{A}}_2\left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1\right) < 1. \tag{4.4.4}$$

As also noted in [TS12], the stability of the consensus algorithm is dependent not only on the selection of step sizes but also on the estimation exchange topology $\mathbf{A}_0$. This fact limits the usage of consensus algorithm in practice.

For the diffusion LMS based algorithms, the choice of step sizes in the $\overline{\boldsymbol{\mathcal{M}}}$ of the block diagonal matrix $\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)$ should guarantee that the stability condition (4.4.4) holds, given the left stochastic matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ and by considering the proposed Kronecker extensions. It was shown in [Say12, Lemma D.6], that by using the block maximum norm, denoted as $\|.\|_{b,\infty}$, then for the matrix of type $\overline{\mathbf{A}}_2\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1$, it holds that

$$\begin{aligned}\rho\left(\overline{\mathbf{A}}_2\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1\right) &\le \|\overline{\mathbf{A}}_2\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\overline{\mathbf{A}}_1\|_{b,\infty} \\ &\le \|\overline{\mathbf{A}}_2\|_{b,\infty}\|\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\|_{b,\infty}\|\overline{\mathbf{A}}_1\|_{b,\infty} \\ &= \|\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)\|_{b,\infty} \\ &= \rho\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right).\end{aligned} \tag{4.4.5}$$

Since the matrix $\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)$ is diagonal we impose to have that

$$\rho\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) = \max_k |1 - \bar{\mu}_k| < 1, \tag{4.4.6}$$

where the $\bar{\mu}_k$, $k = 1, \ldots, KM^2$ are the diagonal elements of $\overline{\boldsymbol{\mathcal{M}}}$. Thus based on (4.4.6), the sufficient condition for the (4.4.4) to hold (i.e to make the power component in the (4.4.3) to zero) is to select every $\bar{\mu}_k$ in $\overline{\boldsymbol{\mathcal{M}}}$ so that the diagonal matrix $\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)$ is stable - i.e all the eigenvalues of $\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)$ are inside the unit circle. Since $\overline{\boldsymbol{\mathcal{M}}} = \boldsymbol{\mathcal{M}} \otimes \mathbf{I}_{M^2}$, the step size condition (4.4.6) applies for the diagonal elements μ_k of the $K \times K$ diagonal matrix $\boldsymbol{\mathcal{M}}$ directly. Thus for every $k = 1 \ldots K$ we should have

$$0 < \mu_k < 2. \tag{4.4.7}$$

The CR system designer can choose the step size(s) of the nodes (freely) in the range (4.4.7), by taking into account the CR system design considerations (which are however out of the scope of this work). Usually the step sizes are taken quite small to get more precise estimates (and thus better detection performance) i.e $\mu_k \ll 2$, but with the cost of longer convergence time of the adaptive estimations. We illustrate the effect of convergence in Section 4.5.

Next we analyse the convergence condition of the second component on the RHS of (4.4.3). Based on the result of [HJ12, Corollary 5.6.16] the geometric series

$S_n = \sum_{i=0}^{n-1} \left[\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]^i$ is generated by the matrix $\left[\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]$ and converges if for a matrix norm it holds that $\|\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\| < 1$. This condition guarantees that $\left[\mathbf{I} - \left[\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]\right]$ is invertible. Since from (4.4.5) we have $\rho\left(\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right) \leq \| \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \|_{b,\infty} = \rho\left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right)$, then the sufficient condition for the convergence of the series is given by (4.4.6). Hence when the condition (4.4.6) is satisfied, then as $n \to \infty$ the geometric series converges to

$$S_n = \left[\mathbf{I} - \left[\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]\right]^{-1}. \tag{4.4.8}$$

Thus by noting the mean of $\hat{\mathbf{p}}(n)$ in steady state and under both hypothesis H_i, $i = 0, 1$ as $\mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right]$, we have that

$$\begin{aligned} \mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right] = &\left[\mathbf{I} - \left[\overline{\mathbf{A}}_2 \left(\mathbf{I} - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1\right]\right]^{-1} \\ &\times \overline{\mathbf{A}}_2 \overline{\boldsymbol{\mathcal{M}}\mathbf{C}}\, \mathrm{E}\left[\mathbf{d}(n)|H_i\right], \end{aligned} \tag{4.4.9}$$

where the conditional expectations of observations $\mathrm{E}\left[\mathbf{d}(n)|H_i\right]$ are given in the Section 4.4.2.

The steady state result (4.4.9) is asymptotically biased. Let us note, that the mean error (or bias) in steady state is given as

$$\overline{\mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right]} = \|(\mathbf{1}_K \otimes \mathbf{p}^o|H_i) - \mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_i\right]\|^2, \tag{4.4.10}$$

for, $i = 0, 1$, where $\mathbf{p}^o|H_i$ denotes the optimal solution (4.3.10) and $\mathrm{E}\left[\hat{\mathbf{p}}(\infty)|H_1\right]$ follows from (4.4.9). Since the global solution (4.3.10) follows the Pareto model, we refer in this chapter to the generic result [CS13, Th. 3] for characterizing the bias term, such as (4.4.10). The referred theorem determines that under certain conditions (for example when we have the same step-sizes and a doubly-stochastic matrix $\mathbf{A}$), a lower step-size makes the bias term also lower - i.e the estimates are closer to the optimal solution. Thus in practice, when very low step-size values are used, the bias term can be ignored.

Covariance of estimates

Let us denote the conditional covariance of the estimates under the hypothesis H_i, $i = 0, 1$ as $\mathrm{Cov}\left[\hat{\mathbf{p}}(n+1)|H_i\right]$. Similarly let $\mathrm{Cov}\left[\mathbf{d}(n)|H_i\right]$ denote the conditional covariance of the observations.

Proposition 3. By using recursions (4.4.1), (4.4.2), the definition of covariance and by considering the fact that $\hat{\mathbf{p}}(n)|H_i$ is independent of the stacked observation vector $\mathbf{d}(n)|H_i$, it can be shown that the covariance recursion is

$$\begin{aligned} \mathrm{Cov}\left[\hat{\mathbf{p}}(n+1)|H_i\right] = &\overline{\mathbf{A}}_2 \left(\overline{\mathbf{A}}_0 - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_1 \,\mathrm{Cov}\left[\hat{\mathbf{p}}(n)|H_i\right] \\ &\times \overline{\mathbf{A}}_1^T \left(\overline{\mathbf{A}}_0^T - \overline{\boldsymbol{\mathcal{M}}}\right) \overline{\mathbf{A}}_2^T \\ &+ \overline{\mathbf{A}}_2 \overline{\boldsymbol{\mathcal{M}}\mathbf{C}}\, \mathrm{Cov}\left[\mathbf{d}(n)|H_i\right] \overline{\mathbf{C}}^T \overline{\boldsymbol{\mathcal{M}}\mathbf{A}}_2^T. \end{aligned} \tag{4.4.11}$$

where initial estimate of covariance matrix is noted by $Cov\left[\hat{\mathbf{p}}(0)|H_i\right]$, $i = 0, 1$.

The covariance matrix of the observations, $Cov\left[\mathbf{d}(n)|H_i\right]$, is constant over time n and we provide the values in the Section 4.4.2. Note that (4.4.11) is in the form of a discrete time algebraic Riccati equation (DARE). Thus the covariance results in steady state (i.e the solution to DARE), can be found by using standard procedures, such as [KSH00, App. E].

Finally we note, that according to the theory of adaptive filtering it is generically known that a smaller step size causes lower co-variance of an adaptive estimate in steady state [Say08] and this leads to better detection result.

4.4.2 Statistical modeling of adaptive CM estimates

In this section we first find the theoretical moments for the rank one (Hermitian) observations $\mathbf{d}_{R,k}(n)$, which are then transformed to real domain for the spatio-temporal moment recursions of CM estimate $\hat{\mathbf{p}}_k(n)$, described in the previous subsection. Then we describe the statistical modelling of adaptive CM estimates. Thirdly we propose two methods for approximating the adaptive CM estimates by a Wishart distribution.

Moments of rank one observations

First we summarize the generic and known results about the moments of $M \times M$ NCW and CCCW matrices $\hat{\mathbf{R}}_k$, based on [ST99].

When a $M \times M$ matrix $\hat{\mathbf{R}}_k$ follows a NCW distribution with a DoF parameter $\bar{N}$, a noise population covariance matrix $\bar{\mathbf{\Sigma}}_{v,k}$ and a non-centrality matrix $\bar{\mathbf{\Omega}}_k = \left[\bar{\mathbf{\Sigma}}_{v,k}\right]^{-1}\bar{\boldsymbol{T}}_k$, where $\bar{\boldsymbol{T}}_k = \bar{\mathbf{E}}_k\bar{\mathbf{E}}_k^H$ and where the non-zero column k of $M \times N$ mean matrix $\bar{\mathbf{E}}_k$ is $\mathrm{E}\left[\mathbf{y}_k(n)\right]$, i.e $\hat{\mathbf{R}}_k \sim CW_M(\bar{N}, \bar{\mathbf{\Sigma}}_{v,k}, \bar{\mathbf{\Omega}})$, then the first and vectorized second moments are given as

$$\begin{aligned}
\mathrm{E}\left[\hat{\mathbf{R}}_k\right] &= \bar{N}\bar{\mathbf{\Sigma}}_{v,k} + \bar{\boldsymbol{T}}_k, \\
\mathrm{Cov}\left[\mathrm{vec}(\hat{\mathbf{R}}_k)\right] &= (\bar{\mathbf{\Sigma}}_{v,k}^T \otimes \bar{\boldsymbol{T}}_k) + (\bar{\boldsymbol{T}}_k^T \otimes \bar{\mathbf{\Sigma}}_{v,k}) \\
&\quad + \bar{N}(\bar{\mathbf{\Sigma}}_{v,k}^T \otimes \bar{\mathbf{\Sigma}}_{v,k}).
\end{aligned} \tag{4.4.12}$$

As a special case, when the matrix $\hat{\mathbf{R}}_k$ follows a CCCW distribution with a population covariance matrix $\bar{\mathbf{\Sigma}}_k$, i.e $\hat{\mathbf{R}}_k \sim CW_M\left(\bar{N}, \bar{\mathbf{\Sigma}}_k\right)$, then the matrix $\bar{\boldsymbol{T}}_k$ equals zero and we get

$$\begin{aligned}
\mathrm{E}\left[\hat{\mathbf{R}}_k\right] &= \bar{N}\bar{\mathbf{\Sigma}}_k, \\
\mathrm{Cov}\left[\mathrm{vec}(\hat{\mathbf{R}}_k)\right] &= \bar{N}(\bar{\mathbf{\Sigma}}_k^T \otimes \bar{\mathbf{\Sigma}}_k).
\end{aligned} \tag{4.4.13}$$

These results in [ST99] are based on the characteristic functions of the corresponding Wishart distributions and apply for $\bar{N} \geq 1$. We note that $\bar{\mathbf{\Sigma}}_k^T = \bar{\mathbf{\Sigma}}_k^c$ for a Hermitian matrix and then (4.4.13) also follows from [Ben99] and [WF93]. Thus the moments of $\mathbf{d}_{R,k}(n)$ can be found by using the results (4.4.12) and (4.4.13) with $\bar{N}=1$, $\hat{\mathbf{R}}_k=\mathbf{y}_k(n)\mathbf{y}_k(n)^H=\mathbf{D}_{R,k}(n)$, $\bar{\boldsymbol{\Sigma}}_{v,k}=\sigma_v^2\mathbf{I}_M$, $\bar{\boldsymbol{\Sigma}}_k=\mathbf{R}_{s,k}+\sigma_v^2\mathbf{I}_{M^2}$, $\mathbf{R}_{s,k}=\mathrm{E}\left[|\alpha_k|^2\right]\boldsymbol{\Sigma}_s$ and where in NCW case $\bar{\boldsymbol{T}}_k=\mathrm{E}\left[|\alpha_k|^2\right]\mathbf{m}_s\mathbf{m}_s^H$.

Based on the signal model (4.2.2) and on the AS 1, obviously under H_0 we have that

$$E\left[\mathbf{d}_{R,k}(n)|H_0\right]=\mathrm{vec}\left[\sigma_v^2\mathbf{I}_M\right]. \tag{4.4.14}$$

Under H_1 the mean at node k is given as

$$E\left[\mathbf{d}_{R,k}(n)|H_1\right]=\mathrm{vec}\left[\mathbf{R}_{s,k}+\sigma_v^2\mathbf{I}_M\right]. \tag{4.4.15}$$

Given the network size K, the stacked $KM^2\times 1$ vector $E\left[\mathbf{d}_R(n)|H_i\right]$ over $k=1\dots K$ and for $i=0,1$ can be formed based on the results (4.4.14) and (4.4.15) respectively.

Due to the AS 1, the k,k ($k\in K$) diagonal block of the $KM^2\times KM^2$ network-wise covariance matrix $\mathrm{Cov}\left[\mathbf{d}_R(n)|H_0\right]$ is given as

$$\mathrm{Cov}\left[\mathbf{d}_{R,k}(n)|H_0\right]=\sigma_v^4\mathbf{I}_{M^2}, \tag{4.4.16}$$

while the off-diagonal blocks are zeros, since the observation noise is not correlated over the CR nodes.

The $KM^2\times KM^2$ network-wise $\mathrm{Cov}\left[\mathbf{d}_R(n)|H_1\right]$ is constructed as follows. Firstly, when $\mathbf{m}_s=0$ and $\boldsymbol{\Sigma}_s\neq 0$ (i.e Case 2 type) it can be verified, that the $k,j\in K$ blocks of the $\mathrm{Cov}\left[\mathbf{d}_R(n)|H_1\right]$ are given as

$$\mathrm{Cov}\left[\mathbf{d}_{R(k,j)}(n)|H_1\right]=\begin{cases}\left[\left(\bar{\boldsymbol{\Sigma}}_k\right)^c\otimes\bar{\boldsymbol{\Sigma}}_k\right], & k=j\\ \left[\left(\mathbf{R}_{s,k,j}\right)^c\otimes\mathbf{R}_{s,k,j}\right], & k\neq j\end{cases} \tag{4.4.17}$$

where $\bar{\boldsymbol{\Sigma}}_k=\mathrm{E}\left[|\alpha_k|^2\right]\boldsymbol{\Sigma}_s+\sigma_v^2\mathbf{I}_{M^2}$ and where for $k\neq j$ $\mathbf{R}_{s,k,j}=\mathrm{E}\left[\mathbf{y}_k(n)\mathbf{y}_j(n)^H\right]=\mathrm{E}\left[\alpha_k\alpha_j^c\right]\boldsymbol{\Sigma}_s$, since due to (AS 1) in this case the observations $\mathbf{y}_k(n)$, $\mathbf{y}_j(n)$ are zero mean Gaussian vectors with independent noise processes. Secondly, when $\mathbf{m}_s\neq 0$ and $\boldsymbol{\Sigma}_s=0$ (i.e Case 1 type) and $k=j$, then the k,k on-diagonal block of $\mathrm{Cov}\left[\mathbf{d}_R(n)|H_1\right]$ is given as

$$\begin{aligned}\mathrm{Cov}\left[\mathbf{d}_{R(k,k)}(n)|H_1\right]=&\left[\left(\sigma_v^2\mathbf{I}_{M^2}\right)^T\otimes\sigma_v^2\mathbf{I}_{M^2}\right]\\&+\left[\left(\mathrm{E}\left[|\alpha_k|^2\right]\mathbf{m}_s\mathbf{m}_s^H\right)^T\otimes\sigma_v^2\mathbf{I}_{M^2}\right]\\&+\left[\left(\sigma_v^2\mathbf{I}_M^2\right)^T\otimes\left(\mathrm{E}\left[|\alpha_k|^2\right]\mathbf{m}_s\mathbf{m}_s^H\right)\right].\end{aligned} \tag{4.4.18}$$

When $k\neq j$, then due to (AS 1) the observation noise is not correlated over the CR nodes and it can be verified, that for the k,j off-diagonal blocks, $\mathrm{Cov}\left[\mathbf{d}_{R(k,j)}(n)|H_1\right]=0$. Given the network size K, the network-wise covariance matrix $\mathrm{Cov}\left[\mathbf{d}_R(n)|H_1\right]$ can be composed by using (4.4.17) and (4.4.18) respectively.

Finally the moments of the real observations (as the inputs for the moment recursions of the estimates $\hat{\mathbf{p}}_k(n)$, provided in the previous subsection) can be given for $i=0,1$ as

$$E\left[\mathbf{d}(n)|H_i\right]=\left[\mathbf{T}^{-1}\otimes\mathbf{I}_{M^2}\right]E\left[\mathbf{d}_R(n)|H_i\right], \tag{4.4.19}$$

and

$$\begin{aligned}\mathrm{Cov}\left[\mathbf{d}(n)|H_i\right] = &\left[\mathbf{T}^{-1} \otimes \mathbf{I}_{M^2}\right] \\ &\times \mathrm{Cov}\left[\mathbf{d}_R(n)|H_i\right]\left[(\mathbf{T}^H)^{-1} \otimes \mathbf{I}_{M^2}\right].\end{aligned} \tag{4.4.20}$$

Distributions of the adaptive estimates

To study the detection performance of the proposed distributed, adaptive LE detector, we need to specify the conditional distributions for the detection test statistics - the LE of

$$\hat{\mathbf{R}}_k(n) = \mathrm{vec}^{-1}\left[\mathbf{T}\hat{\mathbf{p}}_k(n)\right] \tag{4.4.21}$$

under both detection hypothesis. As summarized in (4.3.2), when the estimate $\hat{\mathbf{R}}_k(n)$ is obtained by using the linear, equal weighting based method (4.3.1) in a non-distributed and non-cooperative manner, then according to the definition of Wishart matrices [CD11, Chapter 2], $\mathbf{R}_k(n)$ follows a Wishart distribution. Based on the literature, several results exist for the distributions of the LE of Wishart distributed matrices under both detection hypotheses.

The non-asymptotic cumulative distribution function (CDF) model of the LE of a NCW distributed CM matrix is more complicated for practical and numerical evaluation, compared to the corresponding model of a CCCW distribution. Thus often a NCW distribution is approximated by a CCCW distribution, where the non-centrality part of the NCW distribution is incorporated into the population covariance matrix parameter of the CCCW distribution [GN05, TW12, LWT11].

When the estimate $\hat{\mathbf{R}}_k(n)$ is obtained by using the exponential type of averaging (as used in LMS type of algorithms), then due to different weights at every $n \in N$, it can be seen, that a sum of non-equally weighted Wishart matrices over N is not Wishart distributed [GN05, Theorem 3.3.1, 3.5.2]. Based on (4.4.1) it is easy to verify, that the adaptive CM estimate $\hat{\mathbf{R}}_k(n)$ is an average over non-equally weighted vectorized observation matrices. At iteration step n, at node k the elements of the vectors $\hat{\mathbf{p}}_k(n)$ are weighted equally and fused without changing or mixing the order of the elements of $\hat{\mathbf{p}}_k(n)$. The Hermitian property of the estimated CMs is not affected. Thus we need to seek generic CC(C)W approximations for studying the conditional CDFs of LE of adaptively estimated CMs.

Total and General variance approximations

We propose the usage of two methods for approximating the adaptive CM estimates $\hat{\mathbf{R}}_k(n)$ (4.4.21) by conditional approximate CC(C)W distributions. Thus based on (4.4.13) and we assume that

$$\hat{\mathbf{R}}_k(n)|H_i \sim CW_M\left(\bar{N}_i, \bar{\mathbf{\Sigma}}_{k,i}\right), \tag{4.4.22}$$

for $i = 0, 1$, and where $\sim$ denotes an approximate distribution, $\bar{N}_i$ is the approximating DoF and $\bar{\mathbf{\Sigma}}_{k,i}$ is the approximating population covariance matrix parameter of the corresponding CC(C)W distribution. As shown at next, the values for $\bar{N}_i$ and $\bar{\mathbf{\Sigma}}_{k,i}$ are found by matching the mean and trace or determinant of moments of $\hat{\mathbf{R}}_k(n)|H_i$ with the corresponding moments of the devectorized adaptive estimate $\text{vec}^{-1}\left[\mathbf{T}\hat{\mathbf{p}}_k(n)\right]$ under both detection hypothesis.

Proposition 4. For the approximation (4.4.22), $\bar{\mathbf{\Sigma}}_{k,i}$ is found as

$$\bar{\mathbf{\Sigma}}_{k,i} = \frac{1}{\bar{N}_i}\,\mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right] \tag{4.4.23}$$

and $\bar{N}_i$ can be found using the Total Variance (TV) or General Variance (GV) method, respectively, as

$$\bar{N}_{TV,i} = \left\lceil \frac{\mathrm{Tr}\left[\mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]^c \otimes \mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]\right]}{\mathrm{Tr}\left[\mathbf{T}\,\mathrm{Cov}\left[\mathbf{p}_k(n)|H_i\right]\mathbf{T}^H\right]} \right\rceil \tag{4.4.24}$$

or

$$\bar{N}_{GV,i} = \left\lceil \sqrt[M^2]{\frac{\det\left[\mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]^c \otimes \mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]\right]}{\det\left[\mathbf{T}\,\mathrm{Cov}\left[\mathbf{p}_k(n)|H_i\right]\mathbf{T}^H\right]}} \right\rceil, \tag{4.4.25}$$

where $\mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right] = \text{vec}^{-1}\left[\mathbf{T}\,\mathrm{E}\left[\mathbf{p}_k(n)|H_i\right]\right]$ for $i = 0, 1$.

These results are found as follows. Firstly we insert the $\bar{\mathbf{\Sigma}}_{k,i} = \mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]/\bar{N}_i$ from the first equation of (4.4.13) into the RHS of the second equation of (4.4.13) and we have that

$$\begin{aligned}\mathrm{Cov}\left[\text{vec}(\hat{\mathbf{R}}_k(n)|H_i)\right] = \frac{1}{\bar{N}_i}\Bigg[&\mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]^c \\ &\otimes \mathrm{E}\left[\hat{\mathbf{R}}_k(n)|H_i\right]\Bigg].\end{aligned} \tag{4.4.26}$$

Based on (4.4.2) or (4.4.9) and the first equation of (4.4.13), we equalize the means of matrices $\hat{\mathbf{R}}_k(n)|H_i$ and $\text{vec}^{-1}[\mathbf{T}\mathbf{p}_k(n)|H_i]$ and get (4.4.23). For the DoF, $\bar{N}_i$, to use in the approximation, we adapt the idea proposed in [Zha12, Kha89] and equalize the total variances (i.e the traces of corresponding covariance matrices) of the matrices $\hat{\mathbf{R}}_k(n)|H_i$ and $\text{vec}^{-1}[\mathbf{T}\mathbf{p}_k(n)|H_i]$. Thus based on (4.4.26) we require that $\mathrm{Tr}\left[\mathrm{Cov}\left[\text{vec}(\hat{\mathbf{R}}_k(n)|H_i)\right]\right] = \mathrm{Tr}\left[\mathbf{T}\,\mathrm{Cov}\left[\mathbf{p}_k(n)|H_i\right]\mathbf{T}^H\right]$ for $i = 0, 1$. By solving for $\bar{N}_i$ we have the total variance (TV) type of DoF approximation as given by (4.4.24). An alternative for finding the approximation for $\bar{N}_i$ is to equalize the determinants of both matrices [GN05]. Thus based on (4.4.26), we require that $\det\left[\mathrm{Cov}\left[\text{vec}(\hat{\mathbf{R}}_k(n)|H_i)\right]\right] = \det\left[\mathbf{T}\,\mathrm{Cov}\left[\mathbf{p}_k(n)|H_i\right]\mathbf{T}^H\right]$. Similarly, by solving for $\bar{N}_i$ the general variance (GV) type of DoF approximation is given by (4.4.25).

Obviously the total variance method takes into account only the variances of the elements of the corresponding matrices, while the general variance method includes also the covariances of the elements of the corresponding matrices into the approximation of parameter $\bar{N}_i$. AS observed, by using the proposed TV or GV procedures under hypothesis H_1, a NCW matrix is approximated by the CCCW distribution, by matching the moments of NCW matrix into the CCCW model. This is a desired effect, as we explain in the next section. Based on these results we can proceed with the detection performance analysis.

It can be verified, that under H_0 the DoF value approximations (4.4.24) and (4.4.25) are, via the moment analysis of the adaptive estimate $\mathbf{p}_k(n)$, dependant on the step size parameter μ_k and on the full network topology. Since the same noise power value σ_v^2 is present both in the mean and covariance formulas of the adaptive estimate $\mathbf{p}_k(n)$, then a change in the $\sigma_{v,k}^2$ value does not affect the DoF value under H_0. However under H_1 both the DoF approximations are additionally dependant on the noise power value $\sigma_{v,k}^2$. This effect is illustrated in Section 4.5.

Since under H_0, the DoF parameter does not affect the threshold calculation, then a robust detector can also be applied in Algorithm 4.1, by changing the detection module accordingly. We give an example with the MME detector in Section 4.5. On the other hand, since under H_1 the DoF parameter is affected by the uncertainty in the noise power value, then this effect possibly makes the formula of the theoretical detection performance of a robust detector inaccurate as well, but that robust detector can still be used.

4.4.3 Detection Performance Analysis

In this section we provide formulas for studying the probability of false alarm (P_{FA}) and probability of detection (P_D) of the proposed, adaptive LE detector. For this, we need to evaluate the conditional CDFs of the LE of adaptive CM estimate $\hat{\mathbf{R}}_k(n)$ (4.4.22) under both detection hypotheses and under the assumption that $\hat{\mathbf{R}}_k(n)$ is approximated by a CC(C)W distribution as proposed in Section 4.4.2. The resulting detection performance of LE detector is dependent on the performance of the underlying adaptive, distributed CM estimation. Let the eigenvalues of $\bar{\mathbf{\Sigma}}_{k,i}$ in (4.4.22) be denoted in non-increasing order as $\nu_{1,i} \geq \nu_{2,i} \geq \cdots \geq \nu_{M,i}$.

LE under H_0 Hypothesis

Based on [TW12, Kha64], the $\hat{\mathbf{R}}_k(n)|H_0$ (4.4.22) is assumed to follow the CCW distribution and the eigenvalues of $\bar{\mathbf{\Sigma}}_{k,0}$ are $\nu_{1,0} = \cdots = \nu_{M,0} = \sigma_v^2/\bar{N}_0$. The $P_{FA,e}$, based on the non-asymptotic CDF model of the $\hat{\mathbf{R}}_k(n)|H_0$, is given by

$$\begin{aligned} F_{H_0,e}(x) &= |\det(\hat{\mathbf{A}})| \\ P_{FA,e}(\gamma_{LE,k,e}) &= 1 - F_{H_0,e}(\gamma_{LE,k,e}) \end{aligned} \tag{4.4.27}$$

where the $M \times M$ matrix $\hat{\mathbf{A}}_{i,j} = \binom{\bar{N}_0 - j - i - 1}{i-1} \gamma_R(\bar{N}_0 + i - j, \frac{x}{\nu_{1,0}})$, for $i, j = 1, \ldots, M$ and where $\gamma_R(k, u) = \frac{1}{\Gamma(k)} \int_0^u x^{k-1} e^{-x} dx$ is the regularized incomplete Gamma function. The (ideal) detection threshold $\gamma_{LE,k,e}$, based on the non-asymptotic model is expressed as

$$\gamma_{LE,k,e} = F^{-1}_{H_0,e}(1 - P_{FA,e}) \tag{4.4.28}$$

and can be evaluated in terms of a numerical inversion of the exact CDF formula at a desired $P_{FA,e}$ value. An asymptotic CDF based on the Gaussian approximation of Tracy-Widom distribution is proposed in [TW12]. When $\bar{N}_0 \to \infty$, $M \to \infty$ and $M/\bar{N}_0 \in (0, 1)$, the approximate CDF under H_0 can be given as

$$\begin{aligned}
F_{H_0,g}(x) &= \Phi \left(\frac{x - \mathrm{E}[\lambda_1]|H_0}{\sqrt{\mathrm{Var}[\lambda_1]|H_0}} \right), \\
\mathrm{E}[\lambda_1]|H_0 &= \nu_1 \left(a_{LE} + (b_{LE}(-1.7711)) \right), \\
\mathrm{Var}[\lambda_1]|H_0 &= (\nu_1 b_{LE})^2 (0.8132), \\
a_{LE} &= (\sqrt{M} + \sqrt{\bar{N}_0})^2, \\
b_{LE} &= (\sqrt{M} + \sqrt{\bar{N}_0})(\frac{1}{M} + \frac{1}{\bar{N}_0})^{1/3}.
\end{aligned} \tag{4.4.29}$$

This leads to the $P_{FA,g}$ formula

$$P_{FA,g}(\gamma_{LE,k,e}) = Q \left(\frac{\gamma_{LE,k,g} - \mathrm{E}[\lambda_1]|H_0}{\sqrt{Var[\lambda_1]|H_0}} \right), \tag{4.4.30}$$

where Q is the complementary distribution function of the standard Gaussian and to the threshold formula is

$$\gamma_{LE,k,g} = \mathrm{E}[\lambda_1]|H_0 + \sqrt{\mathrm{Var}[\lambda_1]|H_0}\, Q^{-1}(P_{FA,g}). \tag{4.4.31}$$

As seen in Section 4.4, the calculation of the threshold of the LE detector at node k and time index n requires knowledge of the moments of adaptive CM estimates (present at the reference node k) under hypothesis H_0 i.e $\hat{\mathbf{R}}_k(n)|H_0$. Thus based on the values of step sizes, the noise power, the desired P_{FA}, the provided moment recursions and the distribution parameter approximations models for the $\hat{\mathbf{R}}_k(n)|H_0$ in Section 4.4 can be applied, to evaluate the detection threshold at node k and at time instant n. As seen, CR nodes need to know the noise power value(s) to evaluate the moments of $\hat{\mathbf{R}}_k(n)|H_0$. In practice every node k needs to calculate its own threshold by using the provided procedure. While the threshold at node k can be updated iteratively based on the exact moments of $\hat{\mathbf{R}}_k(n)|H_0$, the steady state moments are preferred in practice.

LE under H_1 Hypothesis

Next we obtain a common model for the non-asymptotic CDF$|H_1$ of the LE of adaptively estimated CM matrix. As explained in Section 4.4.2, we approximate the NCW matrix by a CCCW matrix by matching the moments of the matrices. In Section 4.5 we show this approximation works quite well.

Thus we assume the $\hat{\mathbf{R}}_k(n)|H_1$ is distributed by a CCCW distribution. The CDF of the LE of a CCCW matrix $\hat{\mathbf{R}}_k(n)|H_1$ is given by [ZCW05] as follows

$$F_{H_1,e}(x) = K_{CC} \left| \left\{ \nu_i^{\bar{N}_1 - M + j} \bar{\Gamma}\left(\bar{N}_1 - M + j, \frac{x}{\nu_{i,1}} \right) \right\}_{i,j} \right|,$$

$$K_{CC} = \left[\prod_{i=1}^{M} (\bar{N}_1 - i)! \prod_{j=1}^{M} (M - i)! \right]^{-1} \prod_{k=1}^{M} (k-1)! \tag{4.4.32}$$

for $i, j = 1, \ldots, M$ and where $\bar{\Gamma}(k,u) = \int_0^u x^{k-1} e^{-x} dx$ is the lower incomplete gamma function [GR07, 8.350].

This result follows from [ZCW09, Eq. 1] by integrating the joint PDF of ordered eigenvalues of a CCCW matrix, by using [ZCW09, Corollary 2]. It should be emphasized, that as explained in [ZCW09, Chapter II. B], when some of the eigenvalues of $\bar{\mathbf{\Sigma}}_{k,1}$ are coincident, then [CWS10, Lemma. 2] needs to be used to study the limit [ZCW09, Eq. 3].

However we note, that the direct numerical evaluation of (4.4.32) is complicated and (4.4.32) needs to be simplified due to the possibly large $\bar{N}$ values and large arguments of $\bar{\Gamma}(k,u)$. In case of the matrix dimension is $M = 2$, the eigenvalues of the population covariance matrix are naturally not coincident under H_1 (i.e $\nu_{1,1} > \nu_{2,1}$). It can be shown, that when $M = 2$, the following simplified version of (4.4.32) can be used to evaluate the CDF numerically

$$\begin{aligned} F_{H_1,e}(x) &= \frac{\bar{D}}{\left(\frac{1}{\bar{a}} - \frac{1}{\bar{b}}\right) \bar{a}\bar{b}}, \\ \bar{a} &= \nu_{1,1}\bar{N}_1, \\ \bar{b} &= \nu_{2,1}\bar{N}_1, \\ \bar{D} &= \bar{b}\gamma_R(\bar{N}_1 - 1, \frac{x}{\nu_{1,1}})\gamma_R(\bar{N}_1, \frac{x}{\nu_{2,1}}) \\ &\quad - \bar{a}\gamma_R(\bar{N}_1 - 1, \frac{x}{\nu_{2,1}})\gamma_R(\bar{N}_1, \frac{x}{\nu_{1,1}}), \end{aligned} \tag{4.4.33}$$

where $\gamma_R(k,u)$ is the regularized incomplete gamma function.

Finally the probability of detection of the LE of a CCW matrix under H_1 using the exact CDF model is

$$P_{D,e}(\gamma_{LE,k,e}) = 1 - F_{H_1,e}(\gamma_{LE,k,e}). \tag{4.4.34}$$

As earlier, we observe that the channel gain values and the noise power value are required to complete the chain of approximations for the theoretical detection performance analysis.

4.5 Simulation results

In this numerical simulation Section we investigate the detection performance of the ATC type of distributed, adaptive LE detection algorithm. We describe the exact signal model, used in the simulations and then investigate the probability of false alarm (P_{FA}) and the probability of detection P_D of the proposed algorithms.

4.5.1 Simulation model

The channel gains in the following simulations are assumed to be constant over N and M dimension and are sampled for the CR node $k \in K$ as $\alpha_k \sim CN(0, 1)$. We assume there is only one PU signal present in the CR network i.e $\mathbf{s}(n) = s(n)\mathbf{1}$, where $s(n) \sim CN(0, P_s)$ and $P_s = 1$. Using the same examples as in [SCO13], we use for Case 1: $\mathbf{m}_s = s\mathbf{1}$, $\mathbf{\Sigma}_s = 0$, where s is a complex signal realization, and for Case 2: $\mathbf{m}_s = 0$ and $\mathbf{\Sigma}_s = P_s\mathbf{1}\mathbf{1}^H$. Obviously rank($\mathbf{1}\mathbf{1}^H$)=1. Also in (4.4.17) and (4.4.18) we have $\mathbf{R}_{s,k} = |\alpha_k|^2 P_s\mathbf{1}\mathbf{1}^H$, $\mathbf{R}_{s,k,j} = \alpha_k\alpha_j^c P_s\mathbf{1}\mathbf{1}^H$ and $\bar{\boldsymbol{T}}_k = |\alpha_k|^2 P_s\mathbf{1}\mathbf{1}^H$.

When the CR nodes do not cooperate, the local correlation matrix $\mathbf{R}_k$ (4.2.4) is given as follows

$$\mathbf{R}_k = \left[|\alpha_k|^2 \frac{\mathrm{E}\left[\|\mathbf{s}\|^2\right]}{N}\right]\mathbf{1}\mathbf{1}^H + \sigma_{v,k}^2\mathbf{I}_M. \tag{4.5.1}$$

For Case 1 we assume $|s|^2 = P_s$, where s is a complex signal realization. Then we get $\mathrm{E}\left[\|\mathbf{s}\|^2\right] = NP_s$ for both Case 1 and Case 2. The first moment of the rank one input for these two cases is given as

$$E\left[\mathbf{d}_{R,k}(n)|H_1\right] = \mathrm{vec}\left[|\alpha_k|^2 P_s\mathbf{1}\mathbf{1}^H + \sigma_v^2\mathbf{I}_M\right]. \tag{4.5.2}$$

Network topology selection

To improve the communication link failure resistance in the CR network, but to keep the need for processing the data from neighbor nodes minimal, we propose to select the diffusion topology of the estimates in the CR network, i.e the $\mathbf{A}$ matrix, as a combination of the local ($\mathbf{A}, \mathbf{C} = \mathbf{I}$) and ring-around ($\mathbf{A} = \mathbf{A}_{\text{ring}}^T, \mathbf{C} = \mathbf{I}$) topologies [ATB14b, Eq. 11]. Thus at time instant n, at every node k two $M^2 \times 1$ estimates: the local estimate $\hat{\mathbf{p}}_k(n)$ and the estimate $\hat{\mathbf{p}}_{(k-1)\mathrm{mod}K}(n)$ from node $(k-1)\mathrm{mod}K$ are fused together using equal, constant weight 0.5. Therefore, in the subsequent sections we assume, that $\mathbf{C} = \mathbf{I}$, the matrix $\mathbf{A}$ is in such case doubly stochastic (i.e we have additionally $\mathbf{A}\mathbf{1} = \mathbf{1}$) and all the conditions for selecting elements $a_{l,k}$ and $c_{l,k}$, as listed in the Section 4.3.3, are satisfied.

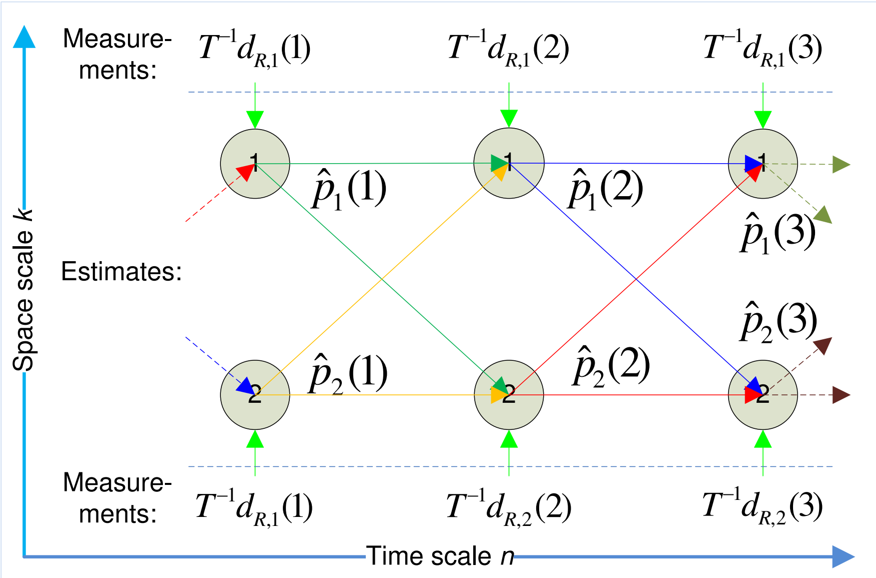

Figure 4.1: Proposed diffusion method

For example when $K = 3$ and by keeping the same notation and conditions for the elements of matrix $\mathbf{A}$, the ring around and diffusion topologies are given as follows

$$\mathbf{A}^T_{\text{ring}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}^T_{\text{diff}} = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix}. \tag{4.5.3}$$

A schematic view of the proposed diffusion and incremental steps for the ATC type of algorithm with $K = 2$ is illustrated in Fig. 4.1.

In the next sections we select the dimension of the estimated matrix as $M = 2$ and use (4.3.6) and (4.4.33). The step size of the algorithms in all the simulations is selected to be $\mu = 0.001$ for all the nodes, unless stated otherwise. Given the step-size value, all the nodes in the network receive $N = 7000$ $[2 \times 1]$ vector-samples to get converged adaptive CM estimates at the last iteration/sample. These CM estimates are used in the simulations to obtain the LE observations. A system designer can choose other values for μ and N (depending on the system requirements).

In Fig. 4.2 we illustrate the change of the LE of adaptively estimated CM with respect to the threshold (4.4.28). We set the noise power to one. After the initialization, the algorithm first tracks and then converges to the steady state level

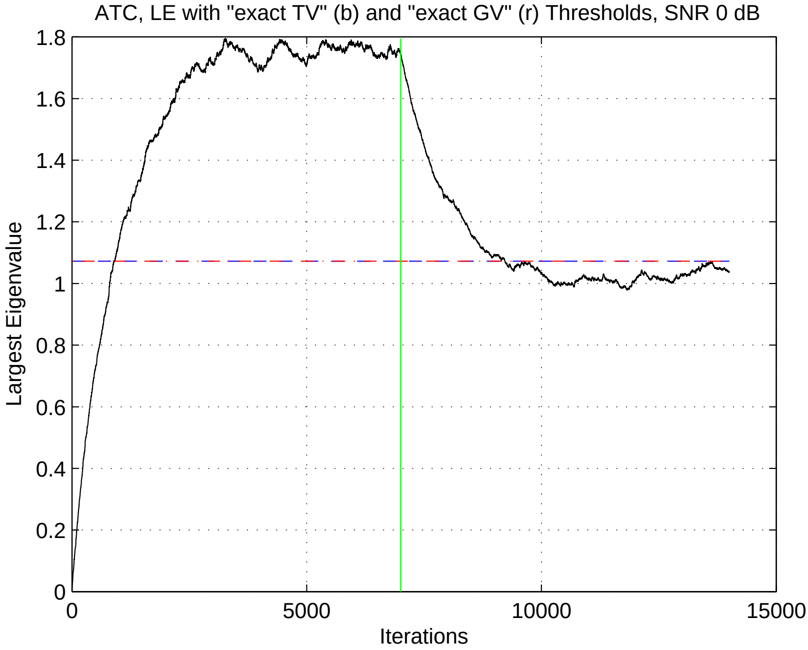

Figure 4.2: LE Adaptive Principle

of LE under the H_1 hypothesis. At time instant 7001 the PU signal switches off, the algorithm adapts and convergences to the H_0 level of the LE value.

DoF values under noise uncertainty

In Fig. 4.3 we illustrate of the effect of the noise power uncertainty to the TV based DoF approximation under Case 2 and H_1. The network sizes are $K = 1, 3, 10, 30$ and the results are taken from the last node in the network. The horizontal axis represents the (network averaged) SNR, which is changed by scaling the noise power value σ_v^2. We use the noise perturbation model [ZL09, Eq. 8] and denote the $\bar{\alpha}$ as the noise uncertainty factor. Two noise value perturbations are added to the non-perturbed case 0 dB ($\bar{\alpha} = 1$): -1 dB ($\bar{\alpha} = 0.796$) and 2 dB ($\bar{\alpha} = 1.585$). As we see, in case of σ_v^2 is inaccurate, then the TV approximated DoF$|H_1$ values are shifted in accordance to the value of $\bar{\alpha}$. For GV based DoF$|H_1$ values, the results are very similar. Also as we already mentioned in Section 4.4.2 that changes, and thus also the pertubations, in $\sigma_{v,k}^2$ do not affect the TV and GV based DoF$|H_0$ approximations. Thus we skip these to latter simulations here.

Next we investigate the performance of the proposed LE algorithms by studying

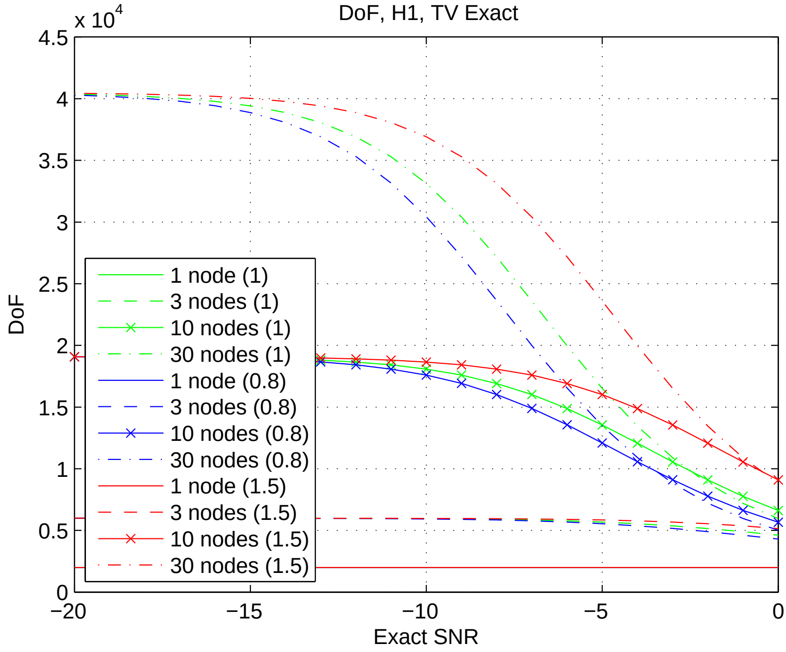

Figure 4.3: ATC, DoF $|H_1$ values with perturbations 0 dB, -1 dB and 2 dB

the P_{FA} in case of PU signal is missing and the P_D, when the PU signal is present. Both the P_{FA} and P_D based on adaptive CM estimates are estimated using the Monte Carlo (MC) method [Kay98]. To have an equal comparison between the node sets in one plot, we take all the reference results from the last node in the network. Obviously, based on the global estimation model (4.3.4), when we have more nodes in the network, then the CM estimates at every node have been better averaged over the channel gain values of the nodes in the CR network.

4.5.2 Probability of false alarm

We start the investigation of the proposed algorithms by studying the P_{FA}. Under the detection hypothesis H_0 we assume $\sigma_v^2 = 1$. We select 21 threshold points in the range of σ_v^2 and determine the LE realizations of adaptive CMs estimates. Then we estimate the P_{FA} over 1000 experiments at every threshold point. The estimated P_{FA} is denoted as **Experiments** in the Fig. 4.4. We compare the estimated P_{FA} with the theoretical P_{FA} models when the Total variance (TV) or the General variance (GV) method are used for determining an approximately equiva-

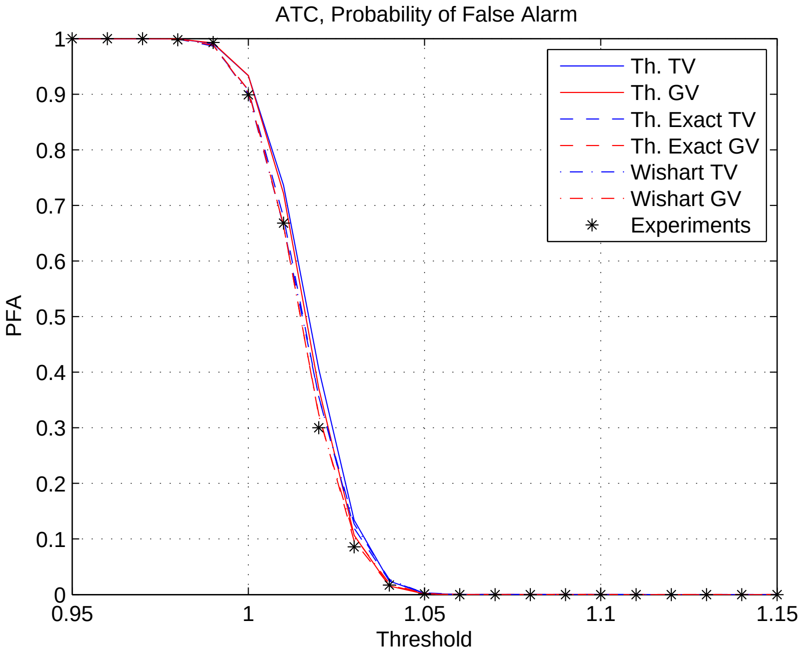

Figure 4.4: P_{FA} versus threshold using ATC

lent CCW matrix. The results using (4.4.29) are denoted as **Th. TV** and **Th. GV** respectively. Similarly the results using (4.4.27) are denoted as **Th. Exact TV** and **Th. Exact GV** respectively. Finally, based on the moments of the adaptive CM estimates, we generate the approximate CCW matrices (by using Cholesky decomposition method), and study the P_{FA} performance based on those matrices in addition (denoted as **Wishart TV** and **Wishart GV** respectively). The P_{FA} versus threshold results are given in Fig. 4.4 for the ATC algorithm. We note that the performance of the TV and GV methods are almost equal and the TV/GV approximations are sufficient for studying the P_{FA} of the adaptive CM estimates. We see a good match between the estimated P_{FA} and the theoretical P_{FA} models are achieved. The Gaussian approximate P_{FA} model (which is easier to use in numerical analyses compared to non-asymptotic P_{FA} model), follows the estimated P_{FA} results quite well and can therefore be used to characterize the P_{FA} of the adaptive estimates. Therefore by knowing the noise power value, the theoretical Gaussian approximate P_{FA} model can be also used for deriving the detection threshold, when we fix a desired P_{FA} value.

4.5.3 Probability of detection

Next we investigate the probability of detection under different noise power conditions using the proposed distributed and adaptive LE detection algorithms with signal models Case 1 or 2. In Case 1 we select one complex PU signal realization, while in Case 2 we set $P_s = 1$ for all the simulations. We note, that the performance of the moment estimation framework of adaptively estimated CMs is well illustrated by the P_D versus SNR analysis. In the P_D/SNR simulations, the change in the (network averaged) SNR is achieved by changing the noise power value σ_v^2. In the comparison of algorithms we use the same individual channel gains of the nodes in all the simulations performed under hypothesis H_1. We set the desired $P_{FA} = 10^{-2}$ for all the nodes. The thresholds of the LE detectors at nodes $k \in K$ are calculated using (4.4.28) with both the TV and GV approximation. Simulations studies showed, that the performance of the Gaussian CDF$|H_1$ based threshold (4.4.31) is almost equal to the performance of the non-asymptotic threshold (4.4.28) and thus not shown in this work.

In the following simulations we compare the performance of 4 different network sizes: $K = 1, 3, 10, 30$ nodes, while the comparable results are taken from the last node in the set. The P_D is estimated over 1000 experiments on a given noise power value. We compare the MC estimated P_D results (based on the adaptively estimated CMs and denoted as **Ad. Exp.** in the figures) with the non-asymptotic theoretical model (4.4.34) (denoted as **Theory**) and with the P_D results based on approximately equivalent CCW matrices (denoted as **W. Exp.**). These latter matrices are generated based on the respective moments under H_1. For the signal model Case 1, the P_D/SNR results are given in Fig. 4.5 and Fig. 4.6 when the TV approximation is used and in Fig. 4.7 and Fig. 4.8 when the GV approximation is used, respectively for the CTA and ATC algorithm. Similarly for the signal model Case 2, the P_D versus SNR results are given in Fig. 4.9 and Fig. 4.10 when TV approximation is used and in Fig. 4.11 and Fig. 4.12 when GV approximation is used, respectively for the CTA and ATC algorithm.

For comparison, the MC estimated P_D/SNR performance of the MME detector [ZL09] under Case 2 is shown additionally in Fig. 4.10 and Fig. 4.12 (where denoted as **MME. Exp.**). The threshold of the MME detector is calculated by using [ZL09, Eq. 29], where in our case $L = 1$ and $N_s = \bar{N}_{TV,0}$ or $N_s = \bar{N}_{GV,0}$. Based on the discussion in Section 4.4.2, it is obvious, that since the noise value perturbations are not affecting the threshold of the MME detector, then the corresponding MC based P_D/SNR performance is not affected as well. In Fig. 4.13 we show a comparison of P_D/SNR performance of the LE detector by using the FC based algorithm in Table 1, TV approximation based exact threshold, and Case 2 model only. In such case the observations of every CR nodes are available for all the CR nodes in the CR network and the CR networks can (independently and adaptively) estimate the CM. In Fig. 4.14 we provide similar comparison of the P_D/SNR performance of the LE detection scheme in Fig. 4.14, by using the consensus algorithm ([TS12]), TV approximation based exact threshold and Case 2 model only and we select

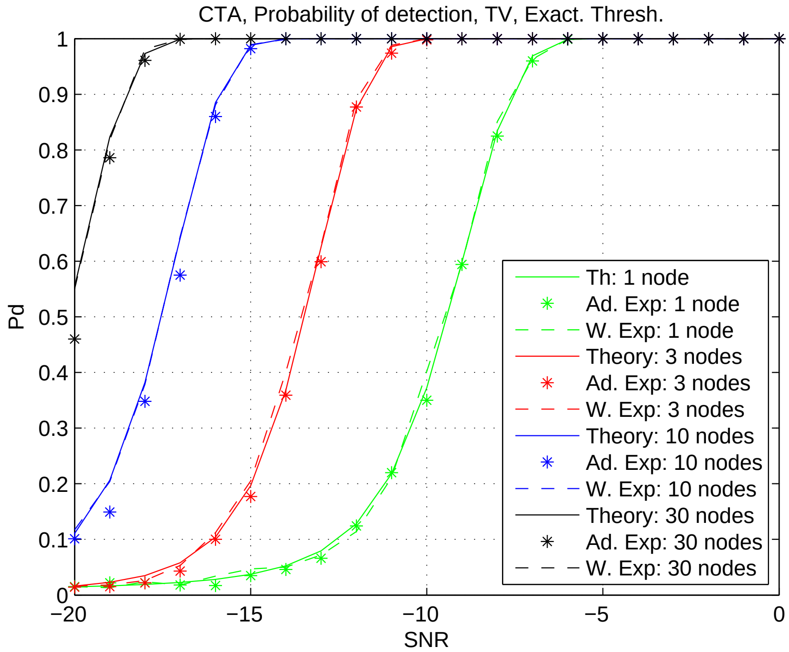

Figure 4.5: Probability of detection, CTA, TV, Case 1

$\mathbf{A}_0 = \mathbf{A}_{\text{diff}}^T$.

We note that the non-asymptotic theoretical P_D model describes the detection performance of adaptively estimated CMs well, also in the low SNR regime. The performance of TV and GV methods is almost equal and thus the TV approximation is computationally less demanding method for the numerical performance analysis of the LE detector. The Case 1 signal model is well approximated by the signal model of Case 2 (CCCW), via the TV and GV based mean and DoF parameter matching.

We observe, that as the number of nodes in the network increases, the point where the P_D starts to decrease from one, moves to the left. In case of one node in the CR network (or in case of the non-cooperating nodes) the P_D is highly dependent on the channel constant of that node. As the number of nodes increases, more channel gain realizations are involved in the network-averaged CM estimation process and thus the P_D results are more equalized over the nodes.

It can be seen, that the LE detector performs better than the MME in terms of perfect detection ($P_D = 1$) in the low SNR region and in case of non-perturbed noise power values.

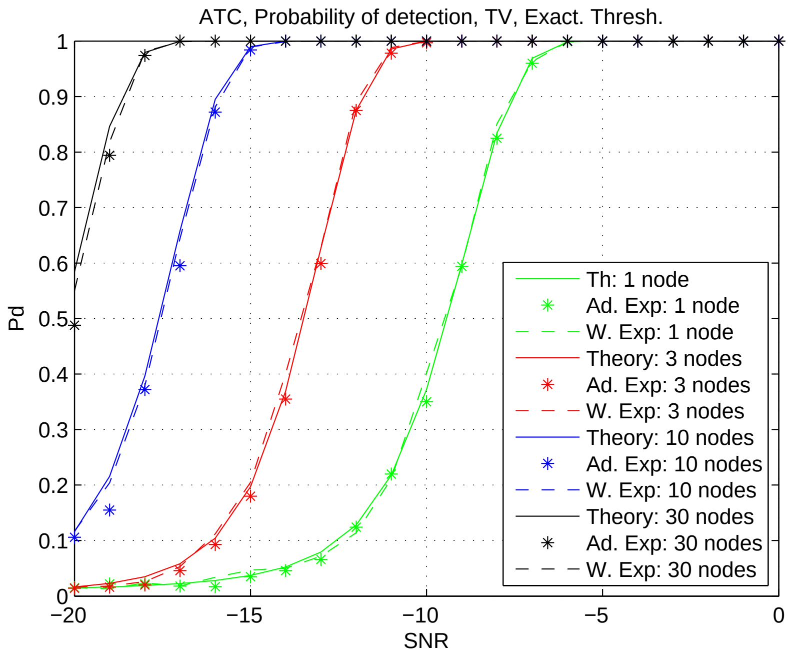

Figure 4.6: Probability of detection, ATC, TV, Case 1

The detection performance of LE detector, when the FC based diffusion LMS algorithm is used, is slightly better, compared with the case of ATC type of LMS. The difference is however not significant. So that in ATC case, where only two exchanges of estimates are allowed for a CR node at time instant n, we can save energy in terms of processing less data at a node k. Also in case of ATC we are not limited to the specific network topology. The detection performance of LE detector, when the consensus algorithm is used, is very similar to the case of the ATC algorithm. As argued in Section 4.4, the usage of ATC type algorithm is less limited by the estimate exchange topology, while this is not the case with the consensus algorithm.

It is clear that the detection performance of the MME detector is not affected by noise power uncertainty also when we use the Diffusion LMS based CM estimation scheme.

Additionally we note that, in [ATB14b] we showed with scalar estimates ($M = 1$) in Case 2, that when there are more nodes in the network, then the ATC performs slightly better compared to the CTA type of algorithm. While ATC fuses more data than CTA [CS10], the difference of detection performance with CTA is rather

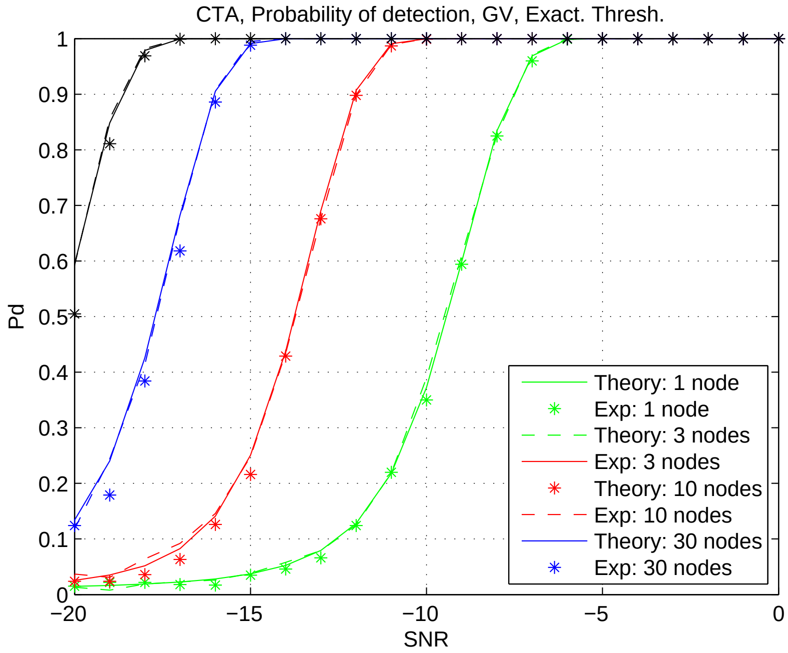

Figure 4.7: Probability of detection, CTA, GV, Case 1

small and thus we also skip these comparisons in this work. We also observed in [ATB14b, ATB14a, ATB14c] that for $K > 30$, P_D does not improve significantly any more.

For illustrating the closeness of the detection results of different CR nodes, we use the theoretical results and plot the P_D/SNR performances of all the CR nodes in the network of size K, in Fig. 4.15, by using the ATC algorithm, the TV based exact threshold and the Case 2 model. The four groups of P_D/SNR results from right to the left in Fig. 4.15 correspond to the network of sizes $K = 1, 3, 10, 30$ accordingly, i.e the leftmost group shows the P_D/SNR results of all the 30 nodes in the CR network. It can be seen, that the detection performances of the CR nodes in the CR network are quite close to each other. In practice we are more concerned about the point, where the P_D starts to decrease from 1. In case of 30 nodes in the network, the deviation slightly increases, but is still sufficiently close.

We observe, that the non-asymptotic CDF models, the TV/GV approximations and CCCW based approximation of NCW type of CMs are usable for studying the performance of the LE detection of adaptively estimated CMs - for determining the threshold and for evaluating the theoretical P_D of the LE detector. When

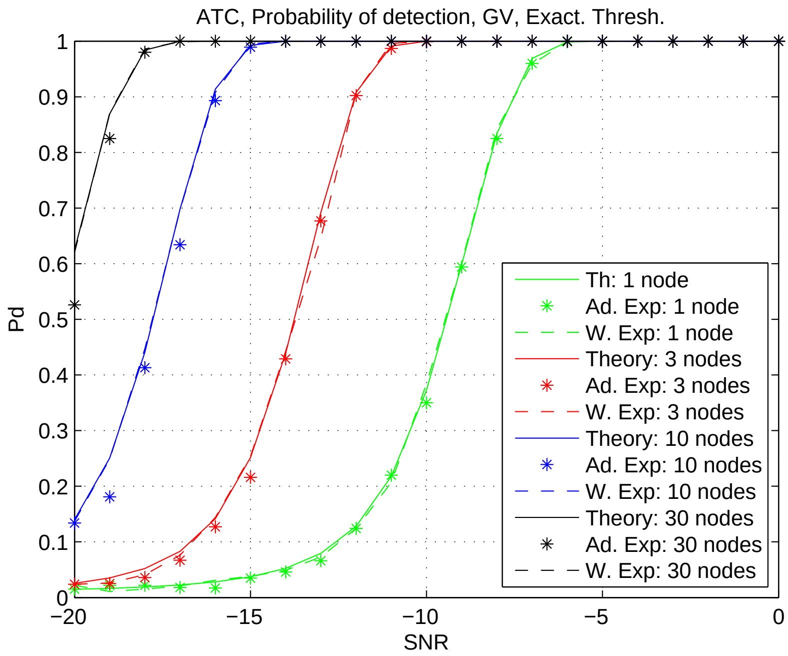

Figure 4.8: Probability of detection, ATC, GV, Case 1

the nodes cooperate in estimating the network-wise CM (while nodes are able to communicate directly only with limited subset of neighbor nodes) then the resulting LE detection performance is equalized and stabilized over the individual CR nodes. We note that other distributed eigenvalue based detection schemes can be studied in similar manner by using the proposed framework in this chapter.

4.6 Conclusion

In this section we studied distributed and adaptive diffusion LMS based LE detection algorithms, which are applicable in CR networks, for detecting the presence of a PU signal. We proposed a network-wise CM estimation model, and derived ATC and CTA type of diffusion based LE detection algorithms. We proposed a general framework for analyzing the performance of the diffusion LMS based LE detection schemes. In our simulation study we demonstrated that the proposed framework and the approximations used for studying the detection performance of the proposed distributed and adaptive LE detection schemes provided matching results between the theory and simulations. The proposed algorithms are able to learn the

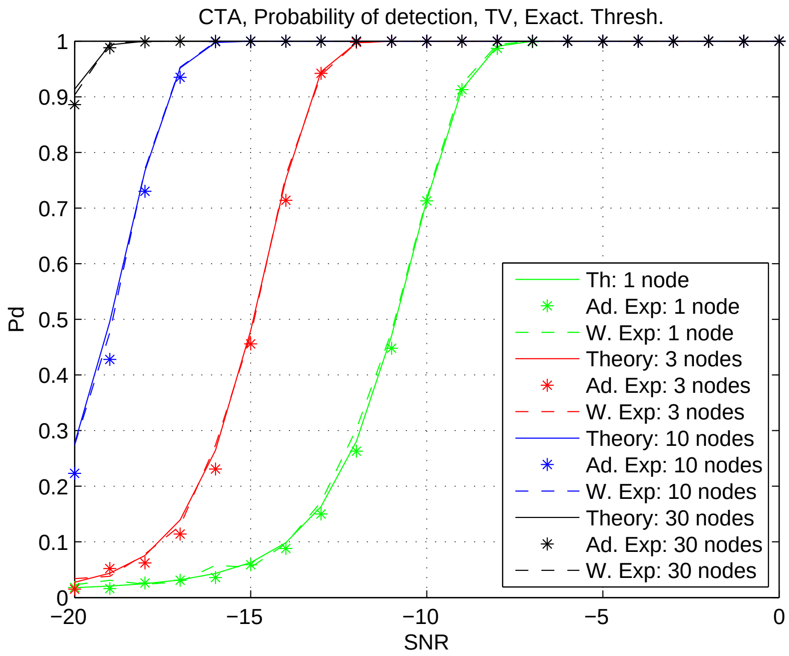

Figure 4.9: Probability of detection, CTA, TV, Case 2

statistical changes in the LE in real time.

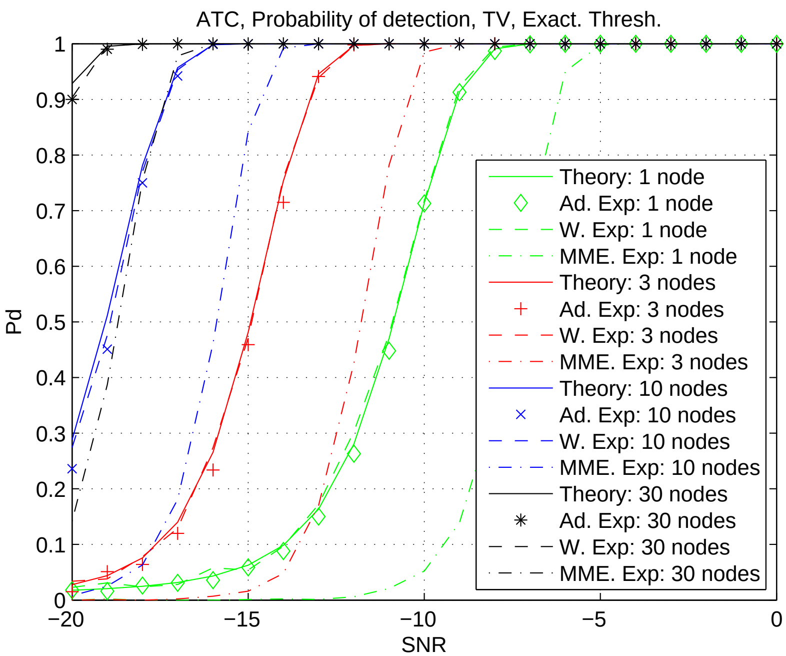

Figure 4.10: Probability of detection, ATC, TV, Case 2

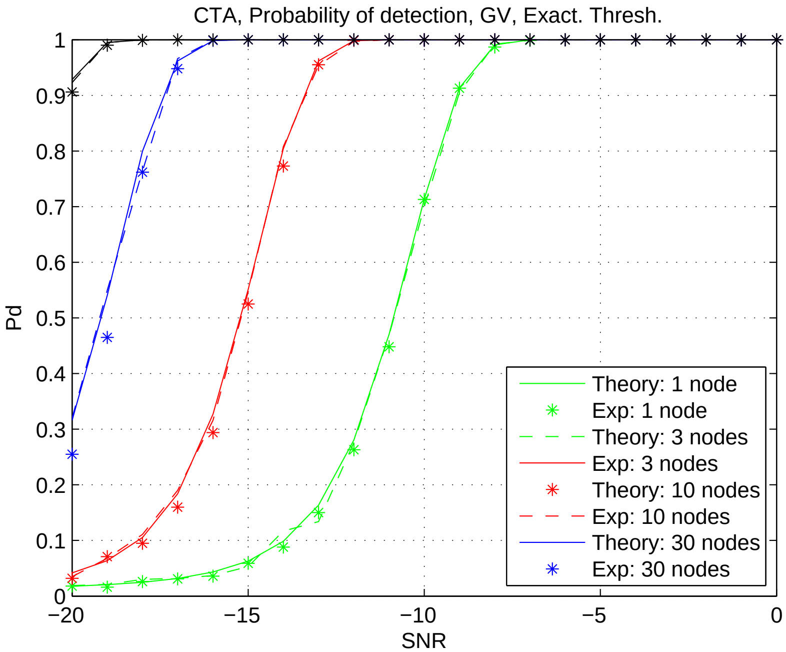

Figure 4.11: Probability of detection, CTA, GV, Case 2

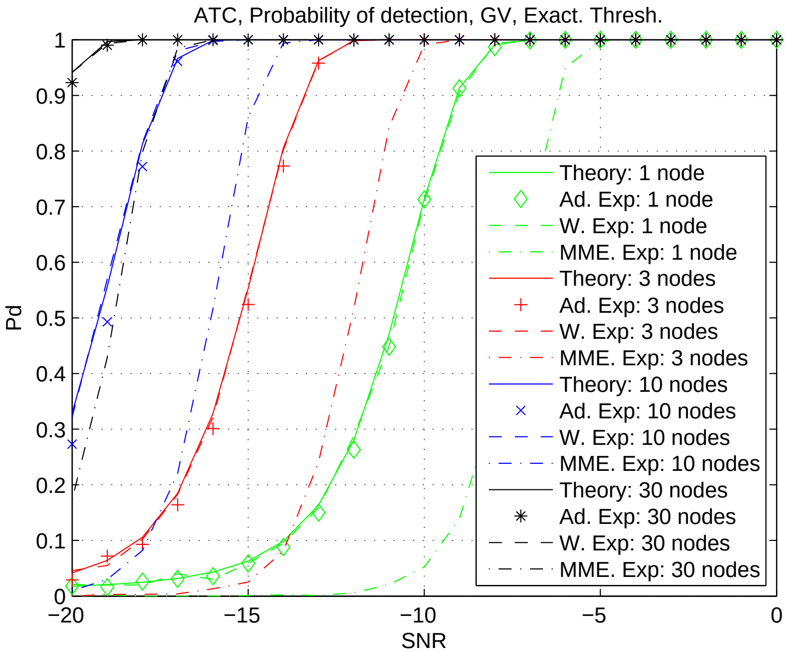

Figure 4.12: Probability of detection, ATC, GV, Case 2

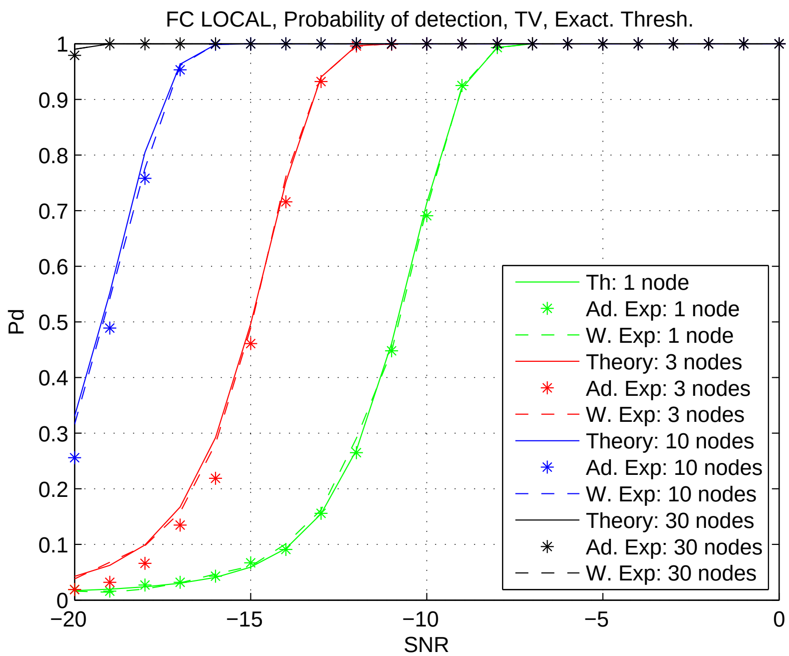

Figure 4.13: Probability of detection, FC, TV, Case 2

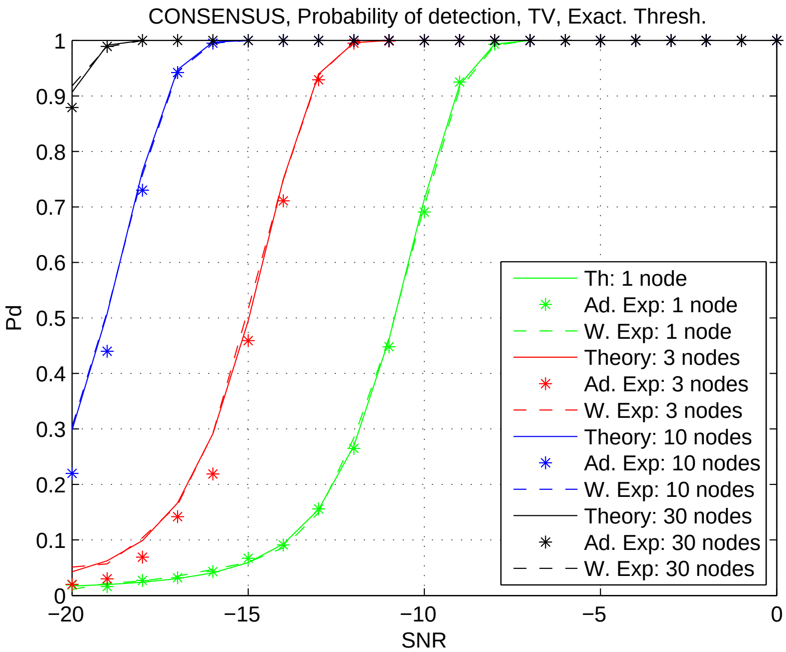

Figure 4.14: Probability of detection, Consensus, TV, Case 2

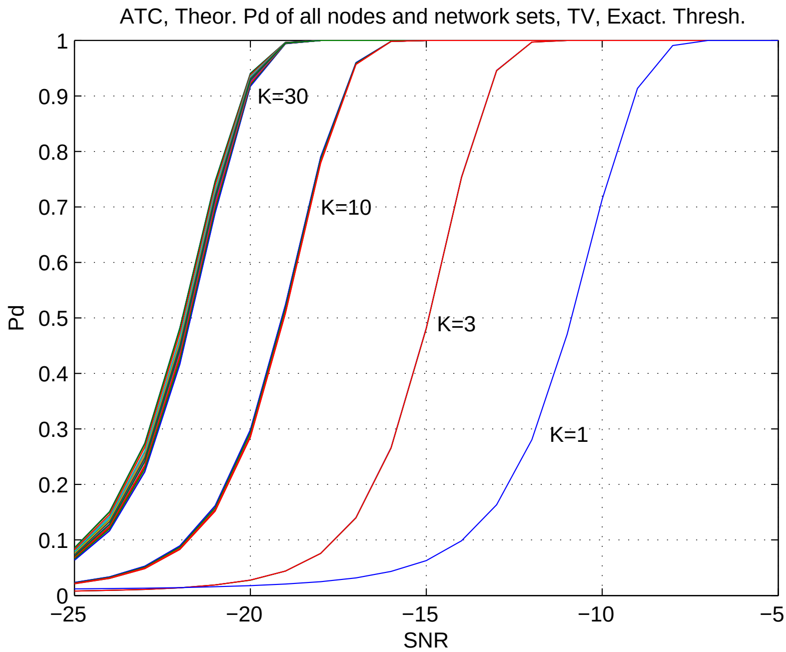

Figure 4.15: Probability of detection, ATC, TV, Case 2, All

Chapter 5

Summary and Future Research

5.1 Summary and Conclusions

This thesis studied two distributed and adaptive detection methods in wireless sensor networks, which are based on a distributed estimation process. The design, implementation and performance study of the proposed algorithms has been done by taking the Cognitive Radio application area into account. In this thesis we studied the algorithms from the estimation and detection domain point of view.

The objective of the current thesis was to design and implement two fully distributed and adaptive detection solutions for Cognitive Radio Networks, namely distributed Energy and distributed Largest Eigenvalue based detection solutions. This objective has been achieved successfully. We have algorithms for each detector to be practically implemented. As common in the research area of estimation and detection, theoretical analytical performance of the proposed solutions was evaluated. We assigned statistical models for the corresponding adaptive estimates and for the detector test statistics to proceed with the moment and detection performance analysis of the proposed algorithms. The theoretical results were verified by computer simulation experiments and good matches between the theoretical performance measures and corresponding experiments were obtained. Thus we proposed and studied two main cooperative, fully distributed and adaptive spectrum sensing methods for a Cognitive Radio Network.

Thematically, the main contributions of the thesis are:

1. **Introduction of an adaptive and fully distributed Energy Detection solution**. We derived and proposed the usage of distributed, diffusion least mean square (LMS) type of power estimation algorithms and three different static network topologies: Ring-Around, Combine and Adapt, Adapt and Combine are studied. The signal power estimation and energy detection solution is not dependant on any Fusion center and the detection decisions can be made independently in every CR nodes or in a selecting a CR node for a network wise decision making. The signal power estimation solution is able to track the changes between the underlying detection hypotheses so that

the usage of such algorithms is more practical in CR network. The detection performance of the proposed schemes was performed by using the statistical properties of these distributed, adaptive estimates. In case of the Ring-Round topology, more specific results about the moment estimation of the distributed estimates were given. With the help of the Central Limit Theorem the distribution of the test statistics of the energy detector was approximated by a CSCG process, by using the moments of the adaptive power estimates. The theoretical findings were verified by MATLAB based simulations. The PU signal, received by a individual CR node may be in deep fading and thus the detection results are dependant on the signal gain value (which is usually unknown for the receiver). We showed that when nodes cooperate in the estimation process of the test statistics, then the resulting detection performance can be significantly improved and stabilized. We also observed that the best detection results (also in terms of lowest variance of the power estimates) are obtained with the ATC type of estimate fusion method and especially when we have about 30 nodes in the network. It was observed that measurement fusion in the diffusion LMS estimation process did not notably improve the resulting detection results.

2. **Introduction of an adaptive and fully distributed Largest Eigenvalue Detection solution**. We selected the Largest Eigenvalue detection method from the domain of correlation matrix based detection methods and designed, implemented an adaptive, fully distributed LE detection solution. Diffusion LMS type of algorithm was implemented with ATC, CTA topologies and with no Fusion Center. In order to study the resulting detection performance we extended the framework of the theoretical performance analysis, from the energy detection solution for the vector estimates. The correlation matrix estimates were vectorized for the distributed adaptive estimation process and after the estimation process re-matrizised. The distribution of the resulting CM matrix estimates was approximated by a Complex Wishard distribution and we implemented the Total and General Variance methods for approximating the Complex Wishart distribution parameters for the mentioned CM distribution approximation. These results were used to proceed with the study of the distribution of the test statistics - the largest eigenvalue of the adaptively estimated CM. The theoretical results were verified by MATLAB based simulations. Similarly we observed that the resulting LE detection performance is more stabilized and equalized over the CR network, when nodes cooperate in the estimation process. Best results were observed with the ATC type of estimate fusion method and there was no notable difference in the performance of the Total and General Variance approximation methods. We justified that the proposed distributed estimation algorithm could be used also with some blind type of detectors, where the noise variance is not needed for the threshold calculation.

Throughout the thesis and in the distributed estimation domain we proposed and used:

3. **Distributed diffusion LMS based scalar and vector estimation algorithms** for estimating the elements of the test statistics of Energy and Largest eigenvalue detector respectively. The algorithms were implemented so that CR nodes jointly participate in the estimation of scalar or vector quantities, where these latter quantities follow the model of network average (to reduce the effect of channel gains), while the CR nodes individually are able to communicate only with a subset of neighbour nodes. Initially the distributed optimization concept for scalar measurements and estimates were derived for the energy detection method. Then a vectorized estimation model of the elements of correlation matrix was proposed. A network topology with minimum number of data fusions in CR network was proposed.

4. **A common framework for the performance analysis of the estimation algorithms and resulting detection performance in CR network**. We derived and proposed the usage of a framework for the performance analysis of the statistical moments of the distributed, adaptive estimates so that several common network topologies and data fusion types are supported. Mean stability analysis for the algorithms was performed. The statistical moments of the distributed estimates were used further in the statistical modelling the test statistics of the selected detectors and then for studying the resulting detection performance.

To conclude, this thesis has shown the benefits of adaptive and fully distributed energy and largest eigenvalue detection solutions for cognitive radio networks. The task of the current thesis, to derive fully distributed versions of two most widely used detectors for cognitive radio, was completed successfully. The obtained results are of practical interest, as the need for opportunistic spectrum sharing in urban areas is increasingly fast.

5.2 Future Research

We now highlight the aspects that might be worthy of future study.

Firstly, in the papers B to E, a constant and equal weighting method for the data fusion between the nodes has been used. Specifically, the matrix $\mathbf{A}$ and $\mathbf{C}$ are taken to be constant and with equal weights over the time instance. In the literature several methods have been proposed for weighting the communication between the nodes in an adaptive manner or for optimizing the combination weights in accordance to a selected optimization criteria. A list of such methods has been given for example in [Say14, Chapter 14]. It has been shown in the literature, that properly optimized weights can improve the properties of the underlying estimation process

− for instance providing better error measures of the estimates. However for instance non−adaptive relative variance rule, as seen in [TS12], would in detection context require knowledge of the channel gains under detection hypothesis H_1 and thus cannot be directly adopted to the detection context. Estimation of these gains would at least require that we know H_1 to be true for a given period of time. The possible implementation of the adaptive combination rule [TS12] in detection context requires an analysis of several additional aspects, which can potentially affect the total detection performance of the main estimation algorithm. It would be interesting to develop an adaptive weight optimization solution, which works also in the detection context - i.e under both detection hypotheses and when we cannot assume that a PU signal is present. When the elements of a data fusion matrix are found based on an additional estimation process and in parallel to the main estimation process, then the fusion weights need to be considered also as random variables. These considerations add an additional complexity to the statistical modelling of the test statistics. The resulting detection performance analysis needs to study also the transition processes of the distributed estimation process − i.e when the algorithm learns new statistical properties, after the detection hypothesis has changed. Tuning of the algorithm parameters needs to be studied. A subspace method based SNR weighting method, as implemented and studied only under H_1 in [UT15], is one of the alternative fusion weight optimization methods, which could be implemented and studied in a detection context.

Secondly, other signal models with various correlation structures and detection methods could be studied together with adaptive and fully distributed estimations methods, other hand diffusion LMS algorithm. The work presented in this thesis did not include comparisons for example with distributed RLS (Recursive Least Squares) method.

Thirdly, it could be interesting to bring in more hardware aspects and constraints to the current research. The current work in this thesis is based on the MATLAB simulations and no major hardware platform specific limitations or aspects have been included so far in our research.

Bibliography

[ALLP12] E. Axell, G. Leus, E.G. Larsson, and H. V. Poor. Spectrum sensing for cognitive radio state-of-the-art and recent advances. *IEEE Signal Processing Magazine*, 29:101–116, May 2012.

[ATB14a] A. Ainomäe, T. Trump, and M. Bengtsson. CTA diffusion based recursive energy detection. In *Latest Trends in Circuits, System Signal Processing and Automatic Control*, pages 38 – 47, Salerno, Italy, 2014.

[ATB14b] A. Ainomäe, T. Trump, and M. Bengtsson. Distributed diffusion LMS based energy detection. In *IEEE 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 176 – 183, St. Petersburg, Russia, 2014.

[ATB14c] A. Ainomäe, T. Trump, and M. Bengtsson. Distributed recursive energy detection. In *IEEE WCNC 2014 Conference Proceedings*, Istanbul, Turkey, 2014.

[Ben99] Mats Bengtsson. *Antenna array signal processing for high rank data models.* PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 1999.

[BGG+13] E. Biglieri, A. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor. *Principles of Cognitive Radio.* Cambridge University Press, Cambridge, 2013.

[Bil95] Patrick Billingsley. *Probability and Measure.* Wiley, New York, 3rd edition, 1995.

[Buc87] K. M. Buckley. Spatial/spectral filtering with linearly constrained minimum variance beamformers. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35:249 – 266, March 1987.

[CD11] R. Couillet and M. Debbah. *Random Matrix Methods for Wireless Communications.* Cambridge University Press, 2011.

[CS10] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, 58:1035–1048, March 2010.

[CS11a] F. S. Cattivelli and A. H. Sayed. Analysis of spatial and incremental LMS processing for distributed estimation. *IEEE Transactions on Signal Processing*, 59:1465–1480, April 2011.

[CS11b] F. S. Cattivelli and A. H. Sayed. Distributed detection over adaptive networks using diffusion adaptation. *IEEE Transactions on Signal Processing*, 59:1917–1932, May 2011.

[CS12] J. Chen and A. H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Transactions on Signal Processing*, 60:4289–305, August 2012.

[CS13] J. Chen and A. H. Sayed. Distributed pareto optimization via diffusion strategies. *IEEE Transactions on Signal Processing*, 7:205–220, April 2013.

[CWS10] M. Chiani, M. Z. Win, and H. Shin. MIMO networks: The effects of interference. *IEEE Transactions on Information Theory*, 56:336–349, January 2010.

[DKM+10] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98:1847–1864, November 2010.

[ER116] Ericsson Mobility Report: On the pulse of the Networked Society. https://www.ericsson.com/assets/local/mobility-report/documents/2016/ericsson-mobility-report-november-2016.pdf, Nov 2016.

[Gar91] W. A. Gardner. Exploitation of spectral redundancy in cyclostationary signals. *IEEE Signal Processing Magazine*, 8:14 – 36, April 1991.

[GJSS16] M. R. Gholami, M. Jansson, E. G. Ström, and A. H. Sayed. Diffusion estimation over cooperative multi-agent networks with missing data. *IEEE Transactions on Signal and Information Processing over Networks*, 2:276–289, September 2016.

[GN00] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, Florida, 1st edition, 2000.

[GN05] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Pearson Prentice Hall, N.J, 1st edition, 2005.

[GR07] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Elviser, Singapore, 7th edition, 2007.

[GSMS09] A. Goldsmith, A. J. Syed, I. Maric, and S. Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97:894 – 914, April 2009.

[Hay02] S. Haykin. *Adaptive Filter Theory, Fourth Edition,.* Prentice Hall, 2002.

[HBS15] M. Hamid, N. Björsell, and B. S. Slimane. Energy and eigenvalue based combined fully blind self adapted spectrum sensing algorithm. *IEEE Transactions on Vehicular Technology*, 65:630 – 642, January 2015.

[HFL+15] L. Huang, J. Fang, K. Liu, H. C. So, and H. Li. An eigenvalue-moment-ratio approach to blind spectrum sensing for cognitive radio under sample-starving environment. *IEEE Transactions on Vehicular Technology*, 64:3465 – 3480, August 2015.

[HJ12] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 2012.

[HNZ09] E. Hossain, D. Niyato, and Han. Z. *Dynamic Spectrum Access and Managament in Cognitive Radio Networks.* Cambridge University Press, 2009.

[HQXZ15] L. Huang, C. Qian, Y. Xiao, and T. Zhang. Performance analysis of volume-based spectrum sensing for cognitive radio. *IEEE Transactions on Wireless Communications*, 14:317 – 330, August 2015.

[HSQ14] L. Huang, H. C. So, and C. Qian. Volume-based method for spectrum sensing. *Digital Signal Processing*, 28:48 – 56, May 2014.

[HWWS14] X. Hong, J. Wang, C. X. Wang, and J. Shi. Cognitive radio in 5g: a perspective on energy-spectral efficiency trade-off. *IEEE Communications Magazine*, 52(7):46–53, July 2014.

[HXZ15] L. Huang, Y. Xiao, and T. Zhang. Robust spectrum sensing for non-circular signal in multiantenna cognitive receivers. *IEEE Transactions on Signal Processing*, 63:498 – 511, January 2015.

[III00] Joseph Mitola III. *An Integrated Agent Architecture for Software Defined Radio.* PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[Kay93] S. M. Kay. *Statistical Signal Processing, Volume I, Estimation Theory.* Prentice Hall, 1993.

[Kay98] S. M. Kay. *Statistical Signal Processing, Volume II, Detection Theory.* Prentice Hall, 1998.

[KGC11] S. Kandeepan, A. Giorgetti, and M. Chiani. Distributed 'ring-around' sequential spectrum sensing for cognitive radio networks. In *Proc. IEEE Int. Conf. on Commun. (ICC'11)*, pages 1–6, Kyoto, Japan, 2011.

[Kha64] C. G. Khatri. Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations. *Annals of the Institute of Statistical Mathematics*, 35:1807–1810, 1964.

[Kha89] C. G. Khatri. Multivariate generalization of t statistic based on the mean square successive difference. *Communications in Statistics - Theory and Methods*, 18:1983–1992, 1989.

[KLW09] S. Kim, J. Lee, and H. Wang. Sensing performance of energy detector with correlated multiple antennas. *IEEE Signal Processing Letters*, 16:671–674, August 2009.

[KSH00] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation.* Prentice Hall, New York, 3rd edition, 2000.

[LCS08] C. G. Lopes, F. Cattivelli, and A. H. Sayed. Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, 56:1865–1877, May 2008.

[LKSS12] J. W. Lee, S. E. Kim, W. J. Song, and A. H. Sayed. Spatio-temporal diffusion strategies for estimation and detection over networks. *IEEE Transactions on Signal Processing*, 60:4017 – 4034, August 2012.

[LS08] C. G. Lopes and A. H. Sayed. Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56:3122–3136, July 2008.

[LWM+17] Gen Li, Tan Wang, Qingyu Miao, Ying Wang, and Biao Huang. *Spectrum Sharing for 5G*, pages 51–73. Springer International Publishing, Cham, 2017.

[LWT11] L. L. Wei and O. Tirkkonen. Approximate condition number distribution of complex non-central correlated wishart matrices. In *Proc. IEEE International Conference of Communications (ICC)*, pages 1 – 5, Kyoto, Japan, 2011.

[LZ09] K. B. Letaief and W. Zhang. Cooperative communications for cognitive radio networks. *Proceedings of the IEEE*, 97:1326–1337, May 2009.

[LZPH08] Y. C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang. Sensing throughput tradeoff for cognitive radio networks. *IEEE Transactions on Wireless Communications*, 7:1326–1337, April 2008.

[MBMS16] V. Matta, B. Braca, S. Marano, and A. H. Sayed. Distributed detection over adaptive networks: Refined asymptotics and the role of connectivity. *IEEE Transactions on Signal and Information Processing over Networks*, 2:442–460, December 2016.

[MM99] J. Mitola and G. Q. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6(4):13–18, Aug 1999.

[MMB16] Constandinos X. Mavromoustakis, George Mastorakis, and Jordi Mongay Batalla. *Internet of Things (IoT) in 5G Mobile Technologies*. Springer Publishing Company, Incorporated, 1st edition, 2016.

[OHMG12] Vicente Osa, Carlos Herranz, Jose F. Monserrat, and Xavier Gelabert. Implementing opportunistic spectrum access in lte-advanced. *EURASIP Journal on Wireless Communications and Networking*, 2012(1):99, 2012.

[PUP02] A. Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[Say08] A. H. Sayed. *Adaptive Filters*. John Wiley and Sons, 2008.

[Say12] A. H. Sayed. Diffusion adaptation over networks. *Academic Press Library in Signal Processing*, 3:323–454, May 2012.

[Say14] A. H. Sayed. Adaptation, learning, and optimization over networks. *Foundations and Trends in Machine Learning*, 7:311–801, July 2014.

[SCO13] S. K. Sharma, S. Chatzinotas, and B. Ottersten. Eigenvalue-based sensing and SNR estimation for cognitive radio in presence of noise correlation. *IEEE Transactions on Vehicular Technology*, 62:3671–3684, October 2013.

[SHT04] A. Sahai, N. Hoven, and R. Tandra. Some fundamental limits in cognitive radio. In *Proceedings of Allerton Conference on Communications, Control and Computing*, 2004.

[SMG09] I. D. Schizas, G. Mateos, and G. B. Giannakis. Distributed LMS for consensus-based in-network adaptive processing. *IEEE Transactions on Signal Processing*, 57:2365–2382, June 2009.

[ST99] S. A. Sultan and D. S. Tracy. Moments of the complex wishart distribution. *International Journal of Mathematical and Statistical Sciences*, 8:51 – 62, June 1999.

[STC+13] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z.J Towfic. Diffusion strategies for adaptation and learning over networks. *IEEE Signal Processing Magazine*, 30:155 – 171, May 2013.

[TNKG10] A. Taherpour, M. Nasiri-Kenari, and S. Gazor. Multiple antenna spectrum sensing in cognitive radios. *IEEE Transactions on Wireless Communications*, 9:814–823, February 2010.

[TS12] S. Y. Tu and A. H. Sayed. Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, 60:6217–2012, December 2012.

[TW12] O. Tirkkonen and L. Wei. Exact and asymptotic analysis of largest eigenvalue based spectrum sensing. *Foundation of Cognitive Radio Systems, InTech DOI: 10.5772/31725, ISBN: 978-953-51-0268-7,*, 16:3–22, March 2012.

[UT15] S. Ulp and T. Trump. Distributed adaptive network with SNR weighed communication. In *IEEE Third International Conference on Digital Information, Networking, and Wireless Communications (DINWC)*, pages 83 – 87, Moscow, Russia, 2015.

[Var96] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1996.

[WF93] A. J. Weiss and B. Friedlander. Performance analysis of spatial smoothing with interpolated arrays. *IEEE Transactions on Signal Processing*, 41:1881 – 1892, May 1993.

[WNK+10] H. Wang, G. Noh, S. Kim, D. Kim, and D. Hong. Advanced sensing techniques of energy detection in cognitive radios. *Journal Of Communications and Networks*, 12:19–29, February 2010.

[WTL14] L. Wei, O. Tirkkonen, and Y. C. Liang. Multi-source signal detection with arbitrary noise covariance. *IEEE Transactions on Signal Processing*, 62:5907–5918, November 2014.

[XBL05] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Fourth International Symposium on Information Processing in Sensor Networks*, pages 63–70, Los Angeles, CA, USA, 2005.

[XBL06] L. Xiao, S. Boyd, and S. Lall. A space-time diffusion scheme for peer-to-peer least-squares estimation. In *The Fifth International Conference on Information Processing in Sensor Networks*, pages 168 – 176, Nashville, TN, USA, 2006.

[Yaj17] Vijaya Yajnanarayana. *Ultra Wideband Communication and Localization*. PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2017.

[YSS13] C. K. Yu, M. Schaar, and A. H. Sayed. Distributed spectrum sensing in the presence of selfish users. In *2013 IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 392–395, St. Martin, France, 2013.

[ZC09] Y. Zeng and Liang Y. C. Spectrum-sensing algorithms for cognitive radio based on statistical covariances. *IEEE Transactions on Vehicular Technology*, 58:1804 – 1815, June 2009.

[ZCW05] A. Zanella, M. Chiani, and M. Win. Performance of MIMO MRC in correlated rayleigh fading environments. In *IEEE Vehicular Technology Conference 2005*, pages 1633 – 1637, Dallas, TX, USA, 2005.

[ZCW09] A. Zanella, M. Chiani, and M. Z. Win. On the marginal distribution of the eigenvalues of wishart matrices. *IEEE Transactions on Communications*, 57:1050–1060, April 2009.

[Zha12] J. Zhang. An approximate hotelling T2 Test for heteroscedastic One-Way MANOVA. *Open Journal of Statistics*, 2:1–11, May 2012.

[ZL09] Y. Zeng and Y. C. Liang. Eigenvalue based spectrum sensing algorithms for cognitive radio. *IEEE Transactions on Communications*, 57:1784–1793, June 2009.

[ZLZ08] Y. H. Zeng, Y. C. Liang, and R. Zhang. Blindly combined energy detection for spectrum sensing in cognitive radio. *IEEE Signal Processing Letters*, 15:649–652, October 2008.

[ZTS12] X. Zhao, S. Y. Tu, and A. H. Sayed. Diffusion adaptation over networks under imperfect information exchange and non-stationary data. *IEEE Transactions on Signal Processing*, 60:3460 – 3475, July 2012.