WIDEBAND COGNITIVE RADIO: SPECTRUM SENSING AND DYNAMIC ACCESS

PH.D THESIS PROPOSAL

SIGNAL THEORY AND COMMUNICATIONS DEPARTMENT
TECHNICAL UNIVERSITY OF CATALONIA
(UNIVERSITAT POLITÈCNICA DE CATALUNYA, UPC)

CANDIDATE:
Josep Font-Segura
Campus Nord D5-214
Jordi Girona 1-3, 08034 Barcelona (Spain)
josep.font-segura@upc.edu

ADVISOR:
Prof. Gregori Vázquez Grau
Campus Nord, D5-204
Jordi Girona 1-3, 08034 Barcelona (Spain)
gregori.vazquez@upc.edu

April 12, 2010

Abstract

Cognitive radio exhibits a tremendous promise for increasing the spectral efficiency for future wireless communication systems. Due to the proliferation of wireless services, the radio spectrum has become severely limited and underutilized in some geographical areas. In cognitive radio networks, new users can opportunistically communicate over the available resources without degrading the licensed primary systems. Two important tasks in cognitive radio are spectrum sensing and dynamic access. On the one hand, spectrum sensing aims at detecting part of the spectrum unused by the primary systems. On the other hand, dynamic access focuses on the methods through which the cognitive users access the resources by employing the information provided by the spectrum sensing detector. Cognitive radio becomes attractive and especially challenging in the wideband regime. As the amount of sensed spectrum increases, the overall occupation by the primary services is low; enabling the recent theory of compressed-sampling to reduce the required sampling rate for detection and dynamic access. This Ph.D. thesis proposal identifies open challenges in wideband cognitive radio research on spectrum sensing and dynamic access, presents a state of the art review on the current literature, and introduces the preliminary results and publications on the field. Based on the state of the art reviews and obtained results, the pending and future research objectives are defined and planned.

Resum

Els sistemes de ràdio cognitiva (cognitive radio) ha esdevingut una tecnologia prometedora per augmentar l'eficiència espectral per als sistemes de comunicacions sense fils del futur. A causa de la proliferació dels serveis sense fils, l'espectre està actualment limitat i desaprofitat in algunes regions. En les xarxes de cognitive radio, els usuaris poden establir comunicació de forma oportuna a través de l'espectre disponible sense degradar les prestacions dels sistemes primaris amb llicència. Hi ha dues tasques primordials en cognitive radio: el sondeig espectral (spectrum sensing) i l'accés dinàmic (dynamic access). Per una banda, spectrum sensing està orientat a detectar les parts de l'espectre que no estan utilitzades pels sistemes primaris. D'altra banda, dynamic access es focalitza en els mètodes d'accés múltiple mitjançant els quals els usuaris cognitius accedeixen a l'espectre, fent servir la informació proporcionada pels detectors d'spectrum sensing. Cognitive radio és especialment atractiva i exigent en amples de banda molt grans (wideband regimes). A mesura que l'ampla de banda a monitoritzar augmenta, l'ocupació per part dels sistemes primaris és baixa, la qual cosa facilita l'aplicació de la teoria de mostreig comprimit (compressed-sampling) per tal de reduir la freqüència de mostreig necessària per a la detecció i l'accés múltiple. Aquesta proposta de tesis identifica varis camps de recerca oberts en spectrum sensing i dynamic access, presenta l'estat de l'art actual, i introdueix els resultats i publicacions obtinguts a la data. Els objectius de recerca pendents i futurs a desenvolupar són definits i planificats a partir de la recerca bibliogràfica i els resultats obtinguts en la temàtica.

The document you are holding in your hands reflects the research done for my Ph.D. thesis since the starting date in September 2009.

The Author

Contents

1 Introduction

In the recent years, wireless services have experimented a tremendous growth, and the demand for wireless access in both voice and high rate data multimedia application has been increasing. New generation wireless communication systems are aimed at accommodating this demand though better resource management and improved transmission technologies.

Cognitive ratio is an attractive, state-of-the-art technology able to cope with this demand which has been motivated by complexity and economy. One the one hand, in wireless networks there has been a trend toward increasingly complex, heterogeneous and dynamic environments. On the other hand, almost all the spectrum resources have been assigned and there is a little new, expensive bandwidth for emerging wireless services and products. In the new paradigm of cognitive radio, the wireless devices employ advanced signal processing techniques together with novel dynamic access methods to support new wireless users who are able to communicate in an opportunistic fashion in the existing crowded spectrum while keeping the existing users from performance degradation.

Most of the traditional communication systems are protected against interferences by means of licensed spectrum allocation for exclusive use. However, with such approach, spectrum resources become underutilized for various reasons. First, the failure of any deployed disruptive technology may lead to unused spectrum. Second, public safety and military radio systems require spectrum for occasional operation, leading to an additional amount of unused spectrum. And third, the technological advances on existing licensed communications diminish the amount of spectrum required for the service. In view of all these observable trends, cognitive radio becomes the solution to a time-varying, flexible use of the spectrum.

The cognitive radio technology is inspired by the Defense Advanced Research Projects Agency (DARPA) Next Generation Communication program, and has been emerged in many other industry and academia research. For example, cognitive radio has already been adopted as core platform in the new standard IEEE 802.22, which is a modification of the IEEE 802.16 and will enable what is referred to as Wireless Regional Area Networks (WRANs), i.e., operating in rural regions in the unused VHF/UHF bands and provide services with performance comparable to that of existing broadband access in urban areas.

This Ph.D. thesis proposal focuses on two main functionalities of the cognitive radio technology: spectrum sensing and dynamic access. The primary function of a cognitive device is to reliably identify available spectrum resources temporally unused by the primary systems in a geographical area —spectrum sensing. Employing this spectral mask, the nodes in a cognitive radio network access the available resources for opportunistic communication (dynamic access). Cognitive radio becomes valuable and especially challenging when the amount of operating spectrum is large —wideband. Wideband regimes are characterized by very low signal-to-noise ratios (SNRs), and very sampling rates are required to enable spectrum sensing and medium access. However, as the sensed spectrum amount increases, it is recognized that the observable signal is scarce in frequency, i.e., the overall spectral occupation is low. The sampling rate of sparse signals can be then dramatically reduced using a novel theory —compressed-sampling. The compressed-sampling theory presumes of sparse signal reconstruction from a compressed set of observations, and has gain recent attention in many fields of the engineering, among them detection.

As a result of the aforementioned premises, the main objective of the present Ph.D. thesis proposal is the design, analysis, and comparison of optimal, efficient signal processing techniques for the problems of spectrum sensing and dynamic access in wideband cognitive radio communications. For this goal, three lines of investigation are identified: optimal spectrum sensing detection based on compressed-sampling second-order statistics, multiple-access techniques for dynamic access, and underlying wide-band cognitive radio communication strategies. The first line of investigation aims at deriving the optimal spectrum sensing algorithms for primary users' activity detection in compressed wideband regimes, and uncovers the main factors that determine the final performance of the spectrum sensing algorithm in realistic scenarios. The second line of investigation focuses on designing multiple-access strategies in wideband cognitive radio networks by making use of the spectral information provided by the spectrum sensing algorithms to perform opportunistic communication without degrading the primary systems. Finally, the third line of investigation merges spectrum sensing and dynamic access in order to conceive efficient and practical communication schemes for cognitive radio in the wideband regime.

The rest of the proposal is organized as follows. Section 2 gives a brief introduction and motivation to the cognitive radio technology, identifies the main open research challenges in the fields of spectrum sensing and dynamic access, and establishes the main objectives of the Ph.D. thesis. The bibliography review on compressed-sampling, spectrum sensing, and dynamic access is provided in Section 3. The preliminary results are introduced in Section 4, as well as the publications in the field. Section 5 identifies the pending and future lines of investigation, and gives the main methodology through which the objectives of the Ph.D. thesis will be achieved..

2 Cognitive Radio

Today's wireless networks are regulated by a fixed spectrum resource assignment, i.e., the spectrum is regulated by governmental agencies and is assigned to license holders or services on a long term basis for large geographical regions. However, a large portion of this assigned spectrum is used only sporadically by the primary services, who concentrate the communication over certain portions of the spectrum while a significant amount of the spectrum remains underutilized. According to recent studies by the U.S. Federal Communications Commission (FCC) [1], the utilization of the available spectrum can be as low as 15% in determined geographical areas.

Specifically out of this premise was born the new paradigm of *cognitive radio*. Cognitive radios utilize advanced signal processing along with novel dynamic spectrum access policies to support new wireless users who opportunistically communicate in the existing congested spectrum without degrading established users. For that purpose, a cognitive radio is an adaptive wireless communications system that takes advantage of all type of side information on the network, e.g., users' activity, channel conditions, modulations and codes format, or even the information message sent by other users who share the spectrum [2]. Depending on the degree of knowledge on this side information, cognitive radio systems are based on underlying, overlaying or interweaving the new users' signals with those of existing users [3].

The *underlay* cognitive radios operate only if the interference caused to the non-cognitive services is below a given threshold. This approach assumes that the secondary users have perfect knowledge of the interference caused by their transmitters to the primary services. One possible approach for transmitting signals causing little interference is to use a large bandwidth over which the cognitive radio signal is spread below the noise level. This is the basis idea of spread spectrum and ultra wideband (UWB) communications. Underlay cognitive radio is the most common scheme when coexisting with licensed primary services, e.g., UWB underlays many licensed bands.

In *overlay* cognitive radio, the secondary users can transmit at any power and simultaneously with the primary users by using sophisticated techniques like dirty paper coding (DPC) or relays. For this to be possible, the cognitive transmitters need to know the primary users' codebooks and messages, which may be impractical in many situations.

Thirdly, the idea of opportunistic communication is exploited in *interweave* cognitive radios, which indeed is the original motivation of cognitive radio [4]. The spectrum gaps or holes can be used sporadically by secondary users for their communication in given time and geographical location conditions. The primal requirement of interweave cognitive radio is that the secondary users must be aware of the spectral activity of the primary systems.

In this Ph.D. thesis proposal, we focus on the interweave concept of cognitive radio and its signal processing and communications challenges. Specifically, we will investigate two of the main functionalities indispensable for the cognitive radio technology to proliferate: spectrum sensing and dynamic access. On the one hand, *spectrum sensing* refers to the signal processing detection techniques that reliably identify the radio resources unused by the primary systems in order to cause little interference. On the other hand, *dynamic access* consists of the enabling communication strategies that capture the best multiple access technique in order to meet the user communications requirements. The envisioned open fields on spectrum sensing and dynamic access are discussed in the sequel.

2.1 Open Challenges in Spectrum Sensing and Dynamic Access

For the last decade, cognitive radio has been one of the major focus of academic research, e.g., [5], as well as application initiatives such as the IEEE 802.22 standard on wireless regional area network (WRAN) [6, 7]. Yet, there are many open research challenges which we identify in the following for both spectrum sensing and dynamic access.

2.1.1 Wideband Spectrum Sensing

Spectrum sensing is an important prerequisite in the envisioned applications of wireless cognitive radio networks. Creating an interference map of the operational spatial region plays a fundamental role in enabling spatial frequency reuse and allowing dynamic spectrum access. Designing fast and reliable spectrum sensing techniques based on cognitive radios local observations is a challenging task with many open research topics.

1. *Sensing duration and sensing rate*.

   The time needed by the spectrum sensing algorithm to detect the presence of primary user activity is probably the most primordial requirement of cognitive radio communications, especially when transmitting because the primary users can claim their spectrum resources at any time. In order to avoid interference from and to the primary users, the secondary users must identify the presence of primary activity as fast as possible and immediately vacate the used band. The main factor that determines a trade-off between sensing time and detection performance is the selection of the parameters to be sensed.

On the other hand, the sensing rate, i.e., how often the cognitive radios perform spectrum sensing, is another important design parameter. The optimum value of the sensing rate depends on the technical capabilities of the cognitive radio and the statistical properties of the primary services [8]. For instance, if the status of the primary users is known to change slowly in time, spectrum sensing can be performed in a more relaxed manner, whereas a more constant spectrum sensing is required in public safety bands to immediately vacate the band to prevent any interference. Selecting the optimum sensing rate is a novel research problem, and the goal is to maximize the cognitive radio channel capacity while maintain the interference to primary users constraint. Many approaches can be proposed to solve these problems. For instance, one could use the guard interval in OFDM systems to perform sensing, or focusing only in the changing parts of the spectrum.

Another problem associated to sensing time is that spectrum sensing cannot be performed on spectrum resources over which the cognitive radio users are communicating. To mitigate this problem, Frequency-Hopping has arisen as a practical solution, as reported in [9].

2. *Detection of spread-spreading primary users.*

   Typically, primary system either consists of a fixed narrowband or spread spectrum services. The two major approaches of spread spectrum communications are the frequency-hopping spread-spectrum (FSSS) and direct-sequence spread-spectrum (DSSS). In both situations, detecting such systems is difficult as the power of the primary users is distributed over a wide frequency range, even though the actual information bandwidth is much narrower [10]. One possible solution is to uncover the hopping pattern or achieve perfect synchronization with the primary users, however such detection in the code-domain is not straightforward.

3. *Wideband sensing.*

   Very high sampling rates are required by conventional spectral estimation methods which have to operate at or above the Nyquist rate when the frequency range to be sensed is very large. Meanwhile, the stringent timing requirements for monitoring the dynamically changing spectrum only allow for a limited number of measurements to be collected for sensing, which makes it challenging to reliably perform high-resolution signal reconstruction. Furthermore, the wideband regime is characterized by close to zero spectral efficiency and SNRs close the minimum required for reliable communication [11]. Therefore, spectrum sensing becomes indeed especially challenging in the wideband regime.

   In this Ph.D. thesis proposal, we discuss the application of the novel theory of compressed-sampling [c.f. 4.1.1] for wideband spectrum sensing. Due to the low overall occupancy of active primary users in large frequency ranges, we recognize that the signals are *sparse* in the spectrum domain. The sampling rate of sparse signals can be then reduced by a large margin and operate below the Nyquist rate. Although there have been many studies on wideband sensing algorithms [c.f. Section 3.1], more researches are still needed especially when the statistical parameters of the primary users' signals are unknown within the frequency range of interest.

4. *Cooperative sensing*.

   Cooperative sensing becomes a practical methodology for increasing the reliability of the cognitive radio network spectrum sensing performance and addresses some problems of individual spectrum sensing such as the hidden primary user problem or punctual deep fading. However, sharing the information among cognitive radios and combining the measurements is a challenging task.

   On the one hand, the shared information can be based on soft-decisions or hard-decisions. Soft-decisions outperform hard-decisions in terms of probability of missed-detection; however they require a larger overhead for transmitting such information.

   On the other hand, there are many approaches for making the final decision. In a distributed network, each node can base its decision on the local measurements of the neighborhood and pass the information in an intelligent manner (e.g., belief propagation). In a centralized network, a decision center collects all the measurements/decisions from the cognitive radios and performs the final decision employing optimal combination (e.g., the likelihood ratio test), or alternatively simpler techniques such as equal gain-combining, selection combining, or switch and stay combining.

5. *Complexity*.

   Finally, complexity is one of the major factors affecting the implementation of practical sensing methods. To detect a signal at very low-SNR and in a harsh environment is not a simple task.

   The research objectives of this Ph.D. thesis proposal focus on statistical-based detection methods. The major advantage of such spectrum sensing detectors is their little dependence on signal and channel knowledge, as well as relatively low complexity. A promising research line is then the design of statistical-based spectrum sensing algorithms that exploit the signal features. At last, detecting the presence of a primary user signal is only the basic task of sensing. For a radio with high level of cognition, further information such as the modulation format employed by the primary service may be exploited. Therefore, signal identification turns to be an advanced task of sensing.

2.1.2 Dynamic Wideband Access

All the cognitive nodes willing to transmit or receive information perform spectrum sensing to acquire a map on the situation of the available spectrum prior to opportunistic communication. Recent theoretical work (see [12] and the references therein for an information-theory formulation of cognitive radio channels) recognizes that the spectral environment is distributed and dynamic. The distributed access occurs due to the fact that the primary activity detected by a cognitive radio differs from that detected around the other cognitive radios.

Multiple-Access Techniques (MATs) are in charge of regulating the access of different users to a common access radio resource, which usually consists of a dedicated frequency bandwidth, but not limited to other natures such as time, code, or space [13]. The common resource in which a set of M users which to communicate is referred as multiple-access channel (MAC) [14]. By its nature, the MAC is interference limited in dense networks. This fact, though, becomes a challenge for cognitive radio MATs, when contrasted to the dedicated MATs, i.e., classical techniques where each user or service has a dedicated (not shared) radio resource. A typical example for such techniques is the time-division multiple-access (TDMA) employed in the GSM mobile network. When considering the MAC, many information-theoretical network issues arise. How do the various senders cooperate with each other to send information to the receivers? What rates of communication are simultaneously achievable? What limitations does interference among the senders and to the primary services put on the total rate of communications?

There are still many open lines of investigation in order to answer all these questions. In the field of dynamic spectrum access [15], the following open challenges are identified.

1. *Sharing dimension*.

   In general, spectrum sharing can be done at any dimension of the spectrum space, i.e., frequency, code, space, and time. Selecting the sharing dimension in cognitive radios is a design parameter that depends on the primary user activity.

   In frequency and code division multiple-access, the main challenge is to perform efficient channel allocation. More advanced signal processing techniques are required in space multiple-access as power control is necessary to avoid interference to the primary systems. Finally, time access is based on the concept of time leases.

2. *Opportunistic Access.*

   Contrary to open spectrum access, i.e., a cognitive radio accessing unlicensed bands, opportunistic dynamic spectrum access is regulated by the priority established between primary and secondary users.

   The primary task in opportunistic access is channel selection. The choice of the frequency bands that can be used for opportunistic communication depends on the amount of primary systems occupancy information available at the cognitive radio receiver. We further distinguish

   (a) *Narrowband Spectrum Access*.

   In Narrowband dynamic spectrum access, the cognitive radios can transmit and receive the information on a predetermined or dynamically chosen frequency band, depending on the primary user occupancy. This gives rise to two techniques: frequency-hopping and frequency-tracking. On the one hand, in frequency-hopping spectrum access, the cognitive transmitter and receiver simultaneously hop across multiple frequencies according to a predetermined sequence. The primal challenge of the frequency-hopping approach is that the transmitter and the receiver need always to be synchronized. On the other hand, frequency-tracking consists of choosing one of the free frequency bands for communication. In this setting, the receiver chooses the expected frequency band based on local observations so as to maximize the probability of coincidence with the transmitter.

   (b) *Wideband Spectrum Access*

In a wideband cognitive radio system, the transmitter and receiver can scan the spectral activity in all the frequency bands. Then, the cognitive radios can span the information using a spanning code over multiple frequency slots that are potentially unused by the primary systems. Unlike the narrowband spectrum access schemes, the secondary users need to have information of all the bands prior to every transmission. The basic challenge in this scheme then goes to the spectrum sensing [c.f. Section 2.1.1].

3. *Centralized or Distributed Spectrum Access*.

   An important aspect of a dynamic spectrum access is whether decisions on the access are taken based on complete or partial information about the current utilization of the spectrum. We then distinguish between the centralized and distributed schemes. In centralized cognitive radio networks, a central device is responsible for gathering the spectrum utilization from all the cognitive radios and performing spectrum allocation and controlling based on the global information on the network. Although the performance of centralized spectrum access is optimal due to the acquisition of network information, reporting to a central device may not be practical in many situations due to communication overhead. Alternatively, in distributed approaches (e.g. [16]) the spectrum access is controlled by each individual cognitive radio based on its own or common policies imposed by the MAT, and neighborhood information. Distributed spectrum access reduces the reporting overhead, but achieves suboptimal performance due to the lack of complete information on the network.

2.2 Research Objectives

Figure 1 presents the framework that encompasses the present Ph.D. thesis proposal goals and work plan. In view on the open research challenges uncovered in Sections 2.1.1 and 2.1.2, the main goal of the Ph.D. thesis entitled ”Wideband Cognitive Radio: Spectrum Sensing and Dynamic Access” is the design, analysis and comparison of optimal, efficient signal processing techniques for the problems of spectrum sensing and dynamic medium access in wideband cognitive radio communications. Recognizing that in wideband communications, the overall combination of the primary systems activity in a given location and time is, by nature, a sparse signal, the research objectives are structured as follows.

1. Optimal spectrum sensing framework for cognitive radio.
2. Information-theoretical analysis of compressed and nonuniform sampling.
3. Spectral analysis of sparse wideband signals.
4. Second-order statistics reconstruction from compressed observations.
5. Compressed spectrum sensing algorithms for wideband cognitive radio.
6. Cooperative spectrum sensing of sparse signals.
7. Traffic capacity asymptotic performance in wideband cognitive radios.
8. Multiple-access techniques for cognitive radio.

9. Communication approaches for underlying cognitive radios.

Objective 1 aims at determining which statistics of the primary users' signal and thermal noise have a more significant implication on the spectrum sensing detection, and which are the sufficient statistics of the observations. In order to introduce the compressed-sampling paradigm in wideband spectrum sensing, Objectives 2 and 3 pretend to analyze from an information-theory and spectral analysis perspective which are the most adequate strategies of compressed-sampling in realistic sparse wideband scenarios. Preliminary results show that the second-order statistics of the observations are a sufficient statistic for spectrum sensing detection based on the generalized likelihood ratio test (GLRT). Hence, the reconstruction of second-order statistics from compressed observations is to be studied in Objective 4. The final performance of optimal spectrum sensing with compressed-sampling will be analyzed in Objective 5, and the use of side information from neighbor cognitive radios will be investigated in Objective 6.

In the field of dynamic medium access, Objective 7 is intended to apply the recent theory of traffic capacity to analyze the fundamental and asymptotic limits in wideband cognitive radio opportunistic access. The design of multiple-access techniques for cognitive radio that exploit the spectrum information provided by the spectrum sensing detection is to be studied in Objective 8. Finally, Objective 9 aims at solving the challenging problem of communication in underlying cognitive radio, i.e., which technology offers reliable data links for the secondary users while keeping the primary systems with little service degradation.

Next, we give a critical review on the recent literature on the fields encompassed by the former objectives, and the preliminary results obtained within some objectives. A more technical, detailed description on the Ph.D. thesis objectives is given in the work plan in Section 5.

3 State of the Art

Based on the identified open challenges in cognitive radio signal processing and communications, the research objectives of this Ph.D. proposal encompass three main issues. These are the compressed-sampling theory, the spectrum sensing, and the dynamic medium access for wideband cognitive radio networks. Within these fields, there have been several recent research efforts which inter-relate the fields in a progressive manner, from compressed-sampling to dynamic spectrum access with gradual overlay among them.

For this purpose, this Section provides a general overview of the fundamentals and new advances on compressed-sampling, a survey on up-to-date spectrum sensing techniques, and an introduction to possible approaches for medium access in wideband cognitive radios.

3.1 Compressed and Non-Uniform Sampling

Conventional approaches to sampling signals follow the well-known Nyquist theorem, in which the sampling rate must be at least twice the maximum frequency present in the information signal. However, the new theory of compressed-sampling asserts that we can recover certain signals from few samples than the Nyquist rate. Compressed-sampling has arisen from the mathematical society and has been applied to many engineering fields. It relays on two fundamental characteristics: the sparsity of

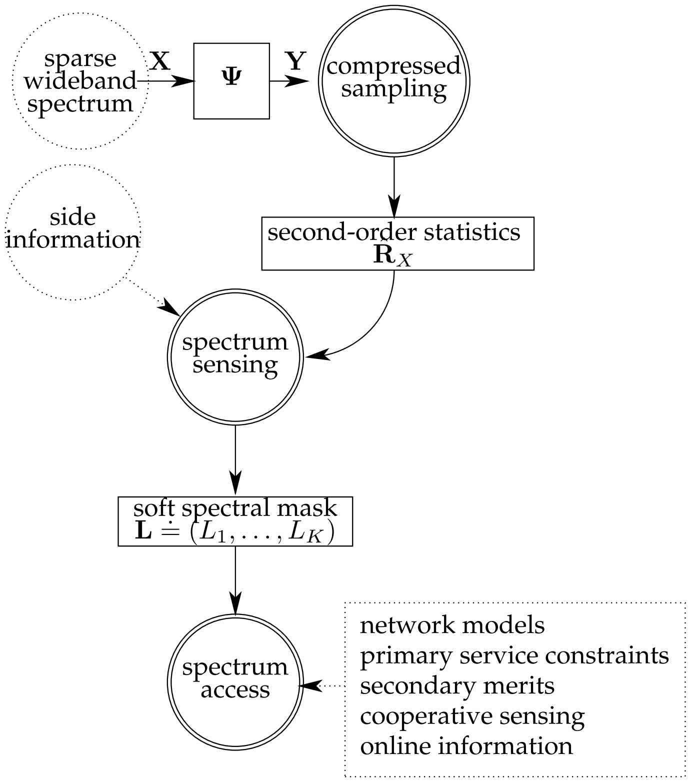

Figure 1: Framework for cross-layer compressed spectrum sensing and access for cognitive radio.

the signal to be sampled, and the incoherence between the sampling method and the underlying domain in which the signal is sparse. The concept of sparsity expresses the idea that the information rate behind a signal can be much smaller than the one suggested by its bandwidth (less degrees of freedom compared to its length).

In the following, we provide the mathematical fundamentals of the classical compressed-sampling theory, and discuss state-of-the-art works which are based or relate to compressed-sampling.

3.1.1 Classical Compressed-Sampling Theory

The compressed-sampling theory has been motivated by E. J. Candès in the last years [17–20]. Consider a continuous-time random process $X(t)$, and an arbitrary N-size sampled vector $\mathbf{x} \doteq (X(t_1), \ldots, X(t_N))^T$, where the sampling intervals t_n accomplish the Nyquist theorem. It is said that $\mathbf{x}$ is *sparse* in the $\mathbf{\Phi}$ domain, if its representation onto the basis $\mathbf{\Phi}$ has only S non-zero coefficients, with $S \ll N$. The compressed-sampling theory claims that $\mathbf{x}$ can be perfectly recovered from the compressed signal $\mathbf{y} = \mathbf{\Psi}(\mathbf{x} + \mathbf{w})$, where $\mathbf{w}$ is the additive Gaussian noise with double-sided spectral density $N_0/2$, and $\mathbf{\Psi}$ is a $K \times N$ compression matrix with enough incoherence with $\mathbf{\Phi}$. The sparsest solution of the coefficients of $\mathbf{x}$ onto $\mathbf{\Phi}$ is given by the ℓ_0-norm constrained minimization problem

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{\ell_0} \tag{1}$$

subject to $\mathbf{y} = \mathbf{\Psi}\mathbf{\Phi}\boldsymbol{\alpha}$, where $\|\boldsymbol{\alpha}\|_{\ell_0}$ denotes the ℓ_0-norm of $\boldsymbol{\alpha}$, which is equal to the number of non-zero elements in $\boldsymbol{\alpha}$, i.e., $\|\boldsymbol{\alpha}\|_{\ell_0} \doteq \lim_{d\to 0} \sum_{n=1}^{N} |\alpha_n|^d$. Solving (1) is a NP-hard problem [17] and very sensitive to the noise. Hence, several alternatives to (1) have been proposed in the literature. Basis pursuit [18] is one of the most used algorithms as replaces the ℓ_0-norm in (1) by the ℓ_1-norm, with a relaxed second-order constraint

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_{\ell_1}, \tag{2}$$

subject to $\|\mathbf{y} - \mathbf{\Psi}\mathbf{\Phi}\alpha\|^2 \geq \epsilon$. The main advantage of basis pursuit is that it leads to linear programming, so (2) can efficiently solved in polynomial time. However, due to the performance degradation in noisy environments, several improvements on the basis pursuit algorithm have been considered, e.g., employing the gradient projection [21], or the interior-point [22] techniques. Alternative options encompass greedy versions of basis pursuit, like the matching pursuit variants [23], which improve the rapidness in solving (2), but may lead to poor estimation performance. Despite the noise sensitivity of the original ℓ_0-norm minimization, it has been shown that it gives the highest possibility of sparse recovery with very few measurements, which motivates the use of approximations to the ℓ_0-norm function to directly solve (1). Among them, the smoothed ℓ_0 cost function [24] approximation is a fast algorithm, and gives significantly improved performance in scenarios with low-SNR regimes. The main idea behind this approach is to approximate the nonlinear function (1) by a suitable continuous function, and take advantage of linearity.

The main drawback of the aforementioned works is that they focus on solving the compressed-sampling problem in a very general fashion, without reaching engineering-like solutions. In the sequel, we discuss novel advances on research related or based on compressed-sampling.

3.1.2 Advances on Compressed-Sampling

Parallel to the Candès work, M. Vetterli *et al.* [25] presented the concept of *r*ate of innovation of signals as a generalization to the concept of bandwidth. The rate of innovation, defined as $\rho \doteq \lim_{\tau\to\infty} \frac{1}{\tau} R_X(\tau)$, where $R_X(\tau)$ counts the number of degrees of freedom of X in an interval τ, expresses the level of innovation per unit of time. The idea behind it is that if a signal has a finite rate of innovation ρ, one hopes to be able to measure and reconstruct the signal by taking only ρ samples per unit of time. There is a clear analogy to compressed-sampling, as ρ may measure the sparsity level of X. However, this generalization allows to consider signals which, even though they are not band limited, they can be sampled uniformly at or above the rate of innovation using appropriate kernels.

From a spectral analysis view point, P. Stoica *et al.* provide a non-uniform sampling approach in [26] and show that non-uniform sampling does not suffer from tome drawbacks present in traditional uniform sampling. The connection to compressed-sampling is feasible due the non-uniformity of the compression (sampling) matrix $\mathbf{\Psi}$.

A third line of investigation related to compressed-sampling has been pursued by G. Eldar *et al.* [27–30]. The work presented in [27,29] establishes a framework related to compressed-sampling, called multi-coset sampling for multi-band signals with sparse support over the sensed band, that establishes the general conditions for perfect reconstruction. This theory is based on universal sampling patters which are a particularization of the compression matrix $\mathbf{\Psi}$ and obey a predetermined structure so that the uniqueness of the solution is preserved. Under some circumstances, if the universal multi-coset pattern takes p samples out of L, it is shown that

$$\frac{1}{T_S} = \frac{p}{LT} \geq nNB,$$

where T_S is the average sampling rate, $1/T$ is the Nyquist rate, and NB is the Landau rate for N bands whose width do not exceed B. If the spectrum support is known, $n = 1$, and the sampling rate must be above the Landau rate. On the other hand, if the spectrum support is known, there is a penalty of $n = 2$ in the minimum required sampling rate. Even the authors provide a reconstruction algorithm based on finite-size observations, numerical algorithms are needed as the reconstruction is based on the classical compressed-sampling theory (1). The theory presented in [27] is applied to wideband analog signals with unknown spectral support in [29]. The idea of multi-coset sampling implementation with filter banks has been further reported in [31,32] for the problem of analog-to-digital conversion.

Sparsity can be naturally found in many signals, and the reconstruction of a compressed sparse signal relies on the incoherence between the compression method and the domain of sparsity. In [30], the authors claim that reconstruction is possible even when the sparsity base is not known, i.e., blind compressed-sampling of *sparke* signals. Uniqueness is ensured for a constrained family of sparsity basis, whose restrictions and reconstruction algorithms are discussed in the work by [30].

Another problem of interest is that of detecting the activity of primary systems employing burst transmissions. Settled in the time-domain, the sparsity does not occur symbol-by-symbol, rather in a block-sparsity fashion. The work in [28] conditions the compressed-sampling problem setting for block-sparse signals, and show that block-sparsity can yield to better reconstruction results than treating the signal as being sparse in the conventional sense.

Finally, a connection to random coding has been established in the recent work by [33]. Random coding [34] is an underlying theory developed by R. G. Gallager [35] for the classical and non-classical, i.e., quantum, channel coding theorem, which presumes of a known a priori distribution of the input alphabet to provide an upper bound on the error probability. The authors in [33] assume that there is some prior distribution on the location of the zero elements of the sparse signal, e.g., which frequency bands are more likely to be empty, and derive information-theoretical bounds on the entropy of the non-zero elements of the sparse signal to show that the compression rate can be further increased.

3.2 Spectrum Sensing

The primary function of a cognitive radio receiver is to reliably identify available spectrum resources temporally unused by primary users in a geographical area. This awareness can be obtained through a database, by using beacons, or by local spectrum sensing at the cognitive radios [36]. In this work, we will focus on spectrum sensing performed at the cognitive radio receivers as it constitutes a broader solution and has less infrastructure requirements. In the sequel, we assume that the cognitive radio receiver has a dataset of N samples of the wideband signal $X(t)$, $\mathbf{X} \doteq (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, and has to decide between hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ whether a primary signal is present in the sensed resource or not, respectively.

3.2.1 Spectrum Sensing Methods for Cognitive Radio

The *energy detector* [37] is the most common way of spectrum sensing because of its low computational and implementation complexities. In addition, it is the most generic detector as the cognitive radio receivers do not need to know any further properties on the primary user signals. The primary user signal is detected by comparing the output of the energy detector, $T = \sum_n \|\mathbf{x}_n\|^2 \geq \lambda$ with a threshold λ that depends on the noise floor. The main challenges behind the energy detector are the threshold computation with inaccurate noise variance, and its poor performance in low-SNR regimes [38]. However, the simplicity of the detector makes its theoretical analysis very clear, and its performance can be accurately characterized.

A second method of spectrum sensing for cognitive radio are the *waveform-based* detectors. Known patterns are commonly employed in wireless systems for synchronization purposes. In the presence of a known pattern pilot signal in the primary services transmission, the cognitive radio can correlate the detected signal with a known copy of itself [38], i.e., $T = \sum_n \mathbf{x}_n^H \mathbf{r}_n \geq \lambda$, where $\mathbf{r}_n$ is the pilot sequence. It is shown that the waveform-based detector outperforms the energy detector in both reliability and convergence time. However, including pattern preambles may not be feasible in some primary services, or it might derive to storage problems when sensing a very wide band.

Cyclostationarity feature detection has been applied in the works by [39, 40] for cognitive radio spectrum sensing. This method exploits the cyclostationarity of the primary users' signals, as periodicity occurs in the signal or in its statistics like the mean of autocorrelation. Even in low-SNR regimes, the cyclostationarity-based detectors can differentiate the primary users' signal from the noise. This is a result of the fact that the thermal noise is a wide-sense stationary with no correlation, while the modulated signals show correlation. The detection is then performed by detecting periodic peaks on the cyclic spectral density function, defined as the Fourier transform of the cyclic autocorrelation $R_X(\tau) \doteq \mathbb{E}[X(t+\tau)X^*(t-\tau)e^{\jmath 2\pi f t}]$. Cyclic frequencies can be assumed to be known or they can be extracted from the actual observations, but inaccurate values on them may cause severe degradation on the performance of the detector.

Finally, *match-filtering* is known as the optimum method for detecting the primary users' signal when the transmitted signal is perfectly known [13]. The main advantage of match-filters is that compared to the aforementioned methods, it achieves high probability of detecting $\mathcal{H}_1$ in a very short time. However, matched-filtering requires the cognitive radios to demodulate the received signals. In other words, it requires perfect knowledge on the primary users' signal deterministic characteristics such as the bandwidth, operating frequency, or modulation type and order. Moreover, its complexity increases as it needs radio-frequency receivers for each signal types. Alternatively, data-independent *filter-banks* detectors have been recently applied to cognitive radio [41] as they provide an integrated tool for wireless communications with less complexity than match-filtering. However, these detectors are not robust to interference. A generalization of filter-banks design has been deployed in [42] in the framework of correlation-matching [43]. The correlation-matching scheme presented in [42] provides power level estimation and frequency location for a candidate correlation matrix to be suited with the sample covariance matrix. If the proper candidate matrix is selected, the resulting detector becomes robust to interferences and significantly outperforms other feature-based detectors.

3.2.2 GLRT-based Spectrum Sensing

The detectors discussed in Section 3.2.1 relay on the fact that the cognitive radio receiver has perfect knowledge on some deterministic parameters of the noise and primary users' signal, such as the noise variance or the signal features, e.g., the cyclic frequency.

In this Ph.D. proposal, we investigate the generalized likelihood ratio test (GLRT) based spectrum sensing detectors for cognitive radio for two reasons. Firstly, the GLRT is the optimal detector in the Neyman-Pearson sense [44] because it maximizes the probability of detection, $\mathbb{P}(\mathcal{H}_1|\mathcal{H}_1)$ in front of a fixed false alarm probability $\mathbb{P}(\mathcal{H}_1|\mathcal{H}_0)$. And secondly, GLRT detectors become a solution to inaccurate parameter models because they natively incorporate a joint parameter estimation framework:

$$L(\mathbf{X}) \doteq \frac{\pi(\hat{\Theta}_1)p(\mathbf{X}|\hat{\Theta}_1, \mathcal{H}_1)}{\pi(\hat{\Theta}_0)p(\mathbf{X}|\hat{\Theta}_0, \mathcal{H}_0)} \geq, \lambda \tag{3}$$

where $\hat{\Theta}_1$ and $\hat{\Theta}_0$ are the maximum a posteriori (MAP) estimates of the unknown model parameters, i.e., $\hat{\Theta}_1 = \arg\max_\Theta \pi(\Theta)p(\mathbf{X}|\Theta, \mathcal{H}_1)$, and $\hat{\Theta}_0 = \arg\max_\Theta \pi(\Theta)p(\mathbf{X}|\Theta, \mathcal{H}_0)$, respectively. If the prior distribution of the unknowns, $\pi(\Theta_1)$ and $\pi(\Theta_0)$ is not available, the GLRT (3) reduces to the classical LRT, and $\hat{\Theta}_1$ and $\hat{\Theta}_0$ becomes the maximum likelihood estimates. GLRT-based cognitive radio detectors have been recently presented [45–47].

A probability-based spectrum sensing scheme for cognitive radio is derived in [45], which discuss a weighted energy detector scheme where each sample is weighted by the probability of the presence of the primary user signal. The main motivation lies in the fact that the primary user signal may appear at any time in between the sensing block, whereas most of the existing literature always assume that the primary user signal appears at the beginning of the block. The probability model establishes that the idle duration of the licensed spectrum band is exponentially distributed.

The authors in [46] derive the optimal GLRT tests for Gaussian signal and noise when all the statistical parameters are unknown, and when the cognitive radio has knowledge on the noise covariance matrix. The main disadvantage of this latter is that an inaccurate value of the noise variance may result in a degradation in performance. The preliminary results in this Ph.D. thesis proposal extend the results in [46] for the wideband low-SNR regime and for unknown noise covariance matrix. We also provide a novel test that predicts the primary user signal level and the noise variance when the cognitive radio has knowledge on the normalized signal covariance matrix, rather than its true value.

The work done in [47] studies the effect of side information on the noise, channel and signal statistics over the performance of the GLRT. The authors also provide extensions to multiple-input multiple-output (MIMO) channels, fast and slow-fading Rayleigh channels, and orthogonal frequency-division multiple-access (OFDMA) schemes. The work concludes that under the white-noise assumption, estimating the noise variance incurs notable performance gain if the true value is inaccurate. This result suggests that the second-order statistics of the noise and signal are a sufficient statistic for the detection problem. GLRT-based spectrum sensing algorithms are also developed in [48] for OFDM cognitive radio networks.

3.2.3 Compressed Spectrum Sensing

A few work has been done in relation to the conjunction of compressed-sampling and spectrum sensing for wideband cognitive radio networks.

Tian *et. al.* [49] study the problem of collaborative distributed spectrum sensing in wideband communications, and provide an extension to spectrum estimation in [50]. The common factor of these works is that the authors employ compressed-sampling and spectrum sensing independently. On the one hand, each cognitive radio reconstructs the received wideband signal from the compressed local observations using the basis-pursiot method, and on the other hand, an energy detector with majority vote decision is performed in the fusion center of the cooperative cognitive radio network. Similarly, the authors in [51] and [52] address the problem of spectrum sensing in wideband cognitive radio scenarios in a two-step approach. Firstly, the basis-pusuit on the Fourier transform of the observations is employed to recover the received spectrum. Secondly, the band location estimation problem is solved using a wavelet-based edge detector.

Alternatively, the work done by Leus *et. al.* [53, 54] on compressed spectrum sensing is focused on the reconstruction of the signal PSD of the primary user signals motivated by the sparsity present on the auto-correlation signal. The authors solve the distributed collaborative sensing problem by considering a fusion center which, after collecting the auto-correlation signals from the cognitive radios, employs the simultaneous orthogonal matching pursuit (SOMP) algorithm to recover the observations PSD. Also, the authors in [55] investigate the problem of cooperative spectrum sensing in cognitive radio networks that show sparsity in two domains: frequency and space. Based on local energy detectors, the cooperative scheme aims to determine the locations of the cognitive transmitters as well as the frequency bands unused by the users for opportunistic frequency reuse. Obtaining an interference map in space and frequency is enforced by statistical models for the location of the cognitive transmitters [56].

All of the aforementioned strategies suffer from distinct drawbacks. First, some establish that the sparsity basis is the Fourier domain in the signal level, rather than the correlation level, which leads to poor performance in communication schemes. Second, the compressed-sampling solution is based on basis-pursuit algorithms, lacking of physical insight interpretation. And third, most of the approaches consist of a two-step effort: once the observations are recovered from the compressed samples, the simple energy detector is applied for detection. In this Ph.D. thesis proposal, we formulate the problem as a joint compressed-sampling and spectrum sensing problem based on second-order statistics (correlation-matching) because while preserving the optimality in the wideband regime, it provides closed-form solutions of statistics estimation which uncovers the fundamental factors that determine the compressed spectrum sensing detector performance.

3.3 Dynamic Spectrum Access

The second primary issue in cognitive radio networks after spectrum sensing is the access strategy that integrates the opportunity exploitation. The main challenge behind multiple-access techniques (MATs) for cognitive radio is the established hierarchical structure in which the secondary system is designed such that no or only insignificant interference is generated toward the primary system.

3.3.1 Multiple-Access Techniques

As a wideband wireless system, the cognitive radio MAT must optimize the utilization of the available spectrum, and guarantee the maximum number of reliable links. Conversely, wideband communications are characterized by arbitrarily large delays, low SNRs, and close to zero spectral efficiency [11]. The spectrum efficiency in classical point-to-point communication channels, defined as the information rate that can be transmitted over a given bandwidth [57], is a suitable metric for benchmarking among different MATs. The spectrum efficiency in wireless networks has recently been casted as the transmission capacity [58, 59]. The transmission capacity is defined as the number of successful transmissions taking place in the network per geographical unit area, subject to a constraint on outage probability. The outage capacity was first presented for both the broadcast channel and the multiple-access channel (MAC) in 1999 [60,61]. The importance of this novel concept lies largely in that it can be exactly derived in some important cases, or tightly bounded in many others, hence providing a useful tool of comparison. In general, we distinguish between the following fundamental multiple-access schemes: time-division multiple-access (TDMA), frequency-division multiple-access (FDMA), code-division multiple-access (CDMA), space-division multiple-access (SDMA), and random access. Spread-spectrum techniques [62] have been adopted in many interference-challenged wireless communication systems because of its efficiency in which several terminals transmit over the same frequency bandwidth, without requiring planned infrastructure (ad hoc). Consequently, CDMA in frequency-hopping (FH-CDMA) and direct-sequence (DS-CDMA) forms are potential dynamic access strategies for cognitive radios.

3.3.2 Transmission Capacity

With no delay constraint, the classical information-theoretical MAC capacity is the relevant performance indicator. This applies, for example, to variable-rate systems. On the other hand, most of today's communication systems carry services for which constant-rate and delay-limited transmission should be considered. In this case, the outage probability is the appropriate capacity metric as it is able to provide the probability that any QoS performance metric such as mutual information, SNR, transmitted power, or bit-error-rate (BER) are bellow or above the resource limit, that is, target transmission rate, target SNR, maximum power or target BER [63–67].

The transmission capacity framework focuses on the statistics of the received signal-to-interference plus noise ratio (SINR) in the MAC. They key underlying this mathematical concept is the use of spatial models for the location of the M terminals on a plane, whose channel response is a function of their relative distances. Spatial models have been used in wireless communications since the late 1970's [68,69], in which the received SINR determines the conditions of transmission success. Moreover, mathematical formulation for spatial models consisting of stochastic geometry has been shown to be well suited for a wide range of problems within wireless communications [70].

The transmission capacity metric was first introduced by the authors in [58]. It is defined as the spatial intensity of attempted transmissions associated to an outage probability of the SINR. The transmission capacity metric has been employed in a wide variety of wireless network design and performance analysis problems, including direct-sequence and frequency-hopping spread spectrum [71], interference-cancellation [72], spectrum sharing of unlicensed, overlaid, and cognitive radio networks [73,74], scheduling and power control [75,76], and the use of multiple antennas for beamforming or orthogonal space-time block coding (OSTBC) [77]. One of the most relevant conclusions of the transmission capacity work is that for CDMA MATs, the ratio between the transmission capacity of FH-CDMA and DS-CDMA only depends on the spreading factor K and the path-loss parameter n by [78]

$$\frac{c_{FH}}{c_{DS}} = K^{1-2/n}.$$

In other words, if $n > 2$, FH-CMDA outperforms DS-CDMA for fixed system conditions by a factor of, e.g. $\sqrt{K}$ in indoor environment with $n = 4$.

Lastly, the work by Menon *et al.* [79], also based on the outage probability and exponential models [80], addresses an assessment on the performance on interference avoidance and interference averaging MATs. The authors show that the interference-avoidance techniques dramatically reduce the interference seen at the primary services. Therefore, underlay schemes such as FH-CDMA should also incorporate interference-avoidance. Moreover, interference-avoidance underlay cognitive radios result in lower outage probabilities as compared to interference-avoidance overlay cognitive radios, with pronounced benefits as the transmission bandwidth is increased.

3.3.3 Practical Opportunistic Spectrum Access Implementations

Not surprisingly, FH-CDMA has already been adopted in many research lines in the field of cognitive radio dynamic spectrum access by L. Tong *et al.* in [81]. In this work, the authors investigate spectrum access policies of cognitive radios in a wireless local area network (WLAN) scenario. The main idea is to sense and predict interference patterns so as to adapt the spectrum access accordingly.

In the line of primary user activity prediction, recent work by L. Tong [82–84] discusses the cognitive medium access problem via stochastic modeling of the primary service. Recognizing the apparent spectrum sensing constraints, it is feasible to assume that a secondary user may not be able to perform full-spectrum sensing or may not be willing to monitor the spectrum when it has no data to transmit. Contrarily, a secondary user can choose to sense a subset of the possible channels. The work in [82] integrates spectrum sensing and spectrum access by adopting a decision-theoretic approach that casts the design of the spectrum access in the framework of Partially Observable Markov Decision Process (POMD) [85]. This formulation leads to optimal policies for spectrum sensing and access, and a systematic trade-off between performance and complexity. Specifically, the traffic statistics of the primary network follow a discrete-time Markov process with 2^K states, being K the number of channels, each state of the type $(S_1(t), \ldots, S_K(t)) \in \mathbb{B}^K$. The hardware and energy limitation on the spectrum sensing capabilities of the secondary users is translated to a maximum number of K_1 sensing and access K_2 channels. The protocol is designed under this constraints, and the spectrum and access decisions are made to maximize the throughput of the secondary user while limiting the interference to the primary network. The authors in [83] further discuss the spectrum access problem in cognitive radio networks with partial and fully-observable radio resources ($K_1 = K_2 = K$). The opportunistic spectrum access via periodic channel sensing is discussed in [84].

Finally, the possibility of spatial reuse in addition to frequency reuse has also received recent growing attention. It was noticed that even if a frequency band is occupied, there could be locations where the transmitted power is low enough so that these frequencies can be reused without suffering from or causing harmful interference to the primary services. These opportunities were investigated in [86] and several statistical models for the location of the transmitters is advocated in [87].

4 Preliminary Results

This Section presents the preliminary results obtained on the topics focused by this Ph.D. thesis proposal. This research includes a novel approach for compressed wideband spectral analysis based on minimum power reconstruction criterion [c.f. Section 4.1], the derivation of optimal GLRT spectrum sensing detectors by exploiting second-order statistics of the primary users' signal and noise [c.f. Section 4.2], the derivation of optimal second-order statistics estimates using compressed observations under a correlation-matching criterion [c.f. Section 4.3], an information-theoretical perspective analysis on the compression matrix [c.f. Section 4.4], and an assessment on multiple-access techniques for cognitive radio networks based on classical information-theory and the novel theory of traffic capacity [c.f. 4.5].

The research efforts performed within this Ph.D. proposal have led to the generation of the following paper list, to which we include former work related to spectrum sensing:

- J. Font-Segura and X. Wang, ”GLRT-Based Spectrum Sensing for Cognitive Radio with Prior Information”, accepted for publication in *IEEE Transactions on Communications*.
- J. Font-Segura and X. Wang, ”Robust Nonlinear Precoding for MU-MIMO-OFDM Systems with Limited Feedback”, submitted to *IEEE Transactions on Wireless Communications*.
- J. Font-Segura, G. Vazquez and J. Riba, ”A Unified Framework for GLRT Spectrum Sensing in Wideband Cognitive Radio”, in preparation for submission to *IEEE Transactions on Communications*.
- J. Font-Segura, G. Vazquez and J. Riba, ”A Compressed Correlation-Matching Approach for Spectral Analysis”, in preparation for submission to *IEEE Transactions on Signal Processing*.

4.1 Compressed Spectral Analysis

As noted in the bibliography review, one of the main problems of the compressed-sampling approaches for wideband spectrum sensing is that the reconstruction algorithms do not give rise to closed-form solutions, hence interpretation insights cannot be easily uncovered. In this work, given the positivity nature of the power spectral density function, we show that sparse spectrum reconstruction is equivalent to a minimum power reconstruction criterion. We derive traditional spectral analysis methods based on the second-order statistics of the compressed observations, rather than on the reconstruction of the data at signal level, which is in general nor guaranteed because the sparsity domain is the spectra. Even this fact, we show that the second-order statistics of the reconstructed signal are equivalent to the reconstructed second-order statistics, i.e., the same estimate on the correlation matrix is achieved.

4.1.1 Sparse Wideband Spectrum Sensing

We consider a wideband signal $X(t)$ which may represent the primary user signal in a cognitive radio network, or the detected signal in an array of sensors. Any sampled realization of the random process X of length N is denoted by the vector $\mathbf{x}_m \doteq [X(t_1^m)\dots X(t_N^m)]$, where the sampling instants t_n^m accomplish the Nyquist rate, i.e., $t_n^m - t_{n-1}^m \leq \frac{1}{2W}$, where W is the sensed bandwidth, for $1 \leq n \leq N$ and $1 \leq m \leq M$. Clearly, in the wideband regime, as $W \to \infty$, in order to recover all the spectral information in the sensed band, we would require an infinite number of samples.

As a stationary random process, the spectrum of X is given as $S_X(f) \doteq \mathcal{F}[R_X(\tau)]$, where $R_X(\tau) \doteq \mathbb{E}[X(t)X^*(t-\tau)]$ is the signal auto-correlation function. In the sequel, we will state that the random process X has a sparse spectrum if the support of $S_X(f)$, i.e., the frequencies such that $S_X(f) > 0$, is small compared to the sensed bandwidth. If X has a spectrum with enough sparsity, we discuss the problem of reconstructing the power spectral density function $S_X(f)$ from a small number of samples, obtained as the linear compression

$$\mathbf{y}_m = \mathbf{\Psi}_m \tilde{\mathbf{x}}_m = \mathbf{\Psi}(\mathbf{x}_m + \mathbf{z}_m), \tag{4}$$

where $\mathbf{z}_m$ is the additive Gaussian noise with double-sided spectral density $N_0/2$, $\tilde{\mathbf{x}}_m$ denotes the noisy observation, and $\mathbf{\Psi}$ is the $K \times N$ compression matrix with compression rate $N/K \geq 1$. In this work, the compression matrices $\mathbf{\Psi}_m$ are built by randomly selecting K rows of the $N \times N$ identity matrix $\mathbf{I}_N$. We will then denote $\mathbf{\Psi}_m$ as a pinning matrix as it randomly selects K samples from the observation $\tilde{\mathbf{x}}_m$. Note that any pinning matrix has the property $\mathbf{\Psi}_m\mathbf{\Psi}_m^H = \mathbf{I}_K$.

Spectrum sensing is aimed at detecting the activity of the primary users over the sensed bandwidth based on the local noisy observations, $\tilde{\mathbf{x}}_m$, for $1 \leq m \leq M$. In a wideband sparse spectrum, the problem is then casted at locating the spectral support of X based on the compressed observations $\mathbf{y}_m = \mathbf{\Psi}_m\tilde{\mathbf{x}}_m$, for $1 \leq m \leq M$. According to the compressed-sampling theory [19,20], the reconstruction of the power spectral density of X is equivalent to solving the following ℓ_1-norm minimization problem

$$\hat{S}_X(f) = \arg \min_{S_X(f)} \|S_X(f)\|_{\ell_1} \tag{5}$$

with the additional constraint $\mathbf{y}_m = \mathbf{\Psi}_m\mathbf{x}_m$ in the noise-free case, or $\|\mathbf{y}_m - \mathbf{\Psi}_m\tilde{\mathbf{x}}_m\|^2 \leq \epsilon$ in the noisy case, both for $1 \leq m \leq M$. However, note that by the positivity nature of the problem, the ℓ_1-norm in

(5) is given by

$$\|S_X(f)\|_{\ell_1} = \int_W S_X(f)df \doteq \mathcal{P}_X = \frac{1}{N}\mathrm{tr}(\mathbf{R}_X), \tag{6}$$

that is, the power of the signal X inside the sensed band of width W, where $\mathbf{R}_X \doteq \mathbb{E}[\mathbf{x}\mathbf{x}^H]$. In other words, we have shown that recovering the sparse spectrum of X obeys a *minimum power reconstruction* (MRP) criterion. We highlight that Equation (6) makes possible the application of many classical spectrum estimators to the compressed version, admitting closed-form solutions for the second-order statistics of X, i.e., $\mathbf{R}_X$ and $S_X(f)$.

4.1.2 Data Reconstruction

We first consider the spectrum analysis problem in the noise-free case, which consists of recover $S_X(f)$ based on the compressed observations $\mathbf{y}_m$ as in (4) with $\tilde{\mathbf{x}}_m = \mathbf{x}_m$. This approach requires the reconstruction of the data, which is not guaranteed as X is not sparse at signal-level. However, since we are not interested in reconstructing the signal, we will show that the correlation of the decompressed signal can be indeed reconstructed.

The periodogram or Blackman-Tuckey spectral estimate is given by [88]

$$\hat{S}_X(f) \doteq \frac{1}{N^2}\mathbf{s}^H(f)\mathbf{R}_X\mathbf{s}(f), \tag{7}$$

where $\mathbf{s}(f)$ is the N-points steering vector $\mathbf{s}(f) \doteq [1\ e^{-j2\pi/N}\ \ldots\ e^{-j2(N-1)\pi/N}]$, and the signal sample covariance matrix is computed as $\mathbf{R}_X \doteq \frac{1}{M}\sum_m \mathbf{x}_m\mathbf{x}_m^H$. The methodology behind data reconstruction consists of computing an estimate of $\mathbf{x}_m$ from the compressed observations $\mathbf{y}_m$, and further determine the sample covariance matrix and the spectrum of X through (7). Using (6), the uncompressed estimate of $(\mathbf{x}_1, \ldots, \mathbf{x}_M)$ is given by the optimization problem

$$(\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_M) = \arg\min_{\mathbf{x}_m} \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{x}_m\mathbf{x}_m^H\right) \quad \text{s.t.} \quad \mathbf{y}_m = \mathbf{\Psi}_m\mathbf{x}_m, \quad 1 \leq m \leq M. \tag{8}$$

The optimization problem (8) is convex on each observation $\mathbf{x}_m$ and the solution is obtained by taking the derivative of the Lagrangian $\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_M) \doteq \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{x}_m\mathbf{x}_m^H\right) + \sum_{m=1}^{M}(\mathbf{y}_m - \mathbf{\Psi}_m\mathbf{x}_m)^H\boldsymbol{\rho}_m$ with respect to $\mathbf{x}_m$. This gives that $\hat{\mathbf{x}}_m = \mathbf{\Psi}_m^H\boldsymbol{\rho}_m$, where the Lagrange multiplier $\boldsymbol{\rho}_m$ is selected to satisfy the constraint in (8), i.e., $\boldsymbol{\rho}_m = (\mathbf{\Psi}_m\mathbf{\Psi}_m^H)^{-1}\mathbf{y}_m = \mathbf{y}_m$, and $\hat{\mathbf{x}}_m = \mathbf{\Psi}_m^H\mathbf{y}_m$. As a result, an estimate of the power spectral density employing the Blackman-Tuckey estimator is given by

$$\hat{S}_X(f) = \frac{1}{N^2}\mathbf{s}^H(f)\mathbf{B}\mathbf{s}(f), \tag{9a}$$

$$\mathbf{B} \doteq \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Psi}_m^H\mathbf{y}_m\mathbf{y}_m^H\mathbf{\Psi}_m. \tag{9b}$$

This indirect method provides an estimate for the sample covariance matrix of the form (9b), i.e., $\hat{\mathbf{R}}_X = \mathbf{B}$, which is sparse by construction. We indicate that for a fixed sampling pattern, i.e., $\mathbf{\Psi}_1 = \ldots = \mathbf{\Psi}_M = \mathbf{\Psi}$, the estimate of the sample covariance matrix becomes $\mathbf{B} = \mathbf{\Psi}^H\mathbf{R}_Y\mathbf{\Psi}$, where $\mathbf{R}_Y$ denotes the sample covariance matrix of the compressed observations, $\mathbf{R}_Y \doteq \frac{1}{M}\sum_m \mathbf{y}_m\mathbf{y}_m^H$. In such a case, we observe that for large data-records the sampling pattern defined by the pinning compression matrix will only recover a fixed subset of K signal modes. Conversely, a variable pinning matrix will randomly select distinct signal modes and hence provide an improved estimate as M increases.

We indicate that in the case of no compression, $\mathbf{B} = \mathbf{R}_X$ and $\hat{S}_X(f)$ become the original sample covariance matrix and the Blackman-Tuckey estimators, respectively.

We next discuss the recovery of the signal power levels according to the traditional Blackman-Tuckey and Capon estimates based on second-order level fitting of the compressed observations. The main motivation is founded by the fact that signal level reconstruction is not guaranteed.

4.1.3 Spectral Amplitude Level Estimate

Recovering the amplitude spectrum of a complex discrete-time signal is important to many applications, such as target feature extraction or radar detection. Consider the following formulation for the original uncompressed data, $\mathbf{x}_m = \alpha(f)\mathbf{s}(f)$, for $1 \leq m \leq M$, where $\alpha(f)$ denotes the complex amplitude of the sinusoidal component with discrete frequency f, and $\mathbf{s}(f)$ is the steering vector. The problem of interest is to recover the power density of X, i.e., $\gamma(f) \doteq |\alpha(f)|^2$, from the compressed data $(\mathbf{y}_1, \ldots, \mathbf{y}_M)$ in (4).

This issue has been addressed in [26] for non-uniformly data sampling, and we further extend this idea for random pinning matrices. In order to recover the spectral power contribution at discrete frequency f, we require the energy fitting from the compressed observations, which is given by the minimization problem

$$\hat{\gamma}(f) = \arg\min_{\gamma(f)} \sum_{m=1}^{M} \|\mathbf{y}_m\mathbf{y}_m^H - \gamma(f)\mathbf{\Psi}_m\mathbf{s}(f)\mathbf{s}^H(f)\mathbf{\Psi}_m^H\|^2, \tag{10}$$

for $1 \leq m \leq M$. The problem (10) is convex on $\gamma(f)$ and is solved by taking the derivative, which is given by $\sum_m \mathrm{tr}(\mathbf{y}_m\mathbf{y}_m^H\mathbf{\Psi}_m\mathbf{s}(f)\mathbf{s}^H(f)\mathbf{\Psi}_m^H) - \gamma(f)\sum_m \mathrm{tr}((\mathbf{\Psi}_m\mathbf{s}(f)\mathbf{s}^H(f)\mathbf{\Psi}_m^H)^2)$. After equaling it to zero and noting that $\mathrm{tr}((\mathbf{\Psi}_m\mathbf{s}(f)\mathbf{s}^H(f)\mathbf{\Psi}_m^H)^2) = K^2$, we obtain that an estimate of the power spectral contribution $\gamma(f)$ is given by

$$\hat{\gamma}(f) = \frac{1}{K^2}\mathbf{s}^H(f)\mathbf{B}\mathbf{s}(f), \tag{11}$$

where $\mathbf{B}$ is expressed as in (9b). According to (11), the spectral power level estimate is of the form of the Blackman-Tuckey spectral estimate where the input correlation matrix is $\mathbf{B}$. By comparing (11) to (9), we observe that the correlation-level fitting optimized in (10) resolves the scaling-factor K^2 in the Blackman-Tuckey estimator, rather than N^2 as the data fitting problem (8). This motivates that the reconstruction of the spectrum of X is possible up to a scaling factor which depends on the compression ratio. In other words, in a compression scenario with normalized compression matrices $\mathrm{tr}(\mathbf{\Psi}_m\mathbf{\Psi}_m^H) = K$, the energy of the observations is not preserved after compression and the adjustment in (11) is required to (9).

We also note that in the case of no compression, (11) reduces to (7).

4.1.4 Capon Power Level Estimate

The main drawback of the power level estimate discussed in Section 4.1.3 is that the information that the signal X is sparse is not incorporated in the formulation.

The Capon power level estimate [89] generalizes the concept of the Blackman-Tuckey estimate by substituting the steering vector by a frequency template $\mathbf{w}(f)$. The design of the frequency template may depend on the purpose of the estimation. According to the MPR criterion deployed in Section 4.1.1, we set the compressed version of the frequency template to minimize the power of the uncompressed observations. The Capon power estimate problem is then given by

$$\hat{\mathbf{w}}(f) = \arg\min_{\mathbf{w}(f)} \hat{\gamma}(f) \tag{12a}$$

$$\hat{\gamma}(f) = \arg\min_{\gamma(f)} \sum_{m=1}^{M} \|\mathbf{y}_m \mathbf{y}_m^H - \gamma(f)\mathbf{\Psi}_m \mathbf{w}(f)\mathbf{w}^H(f)\mathbf{\Psi}_m^H\|^2, \tag{12b}$$

subject to the normalization constraint $\mathbf{w}^H(f)\mathbf{s}(f) = N$. The problem (12b) is convex on $\gamma(f)$ and is solved by taking the derivative with respect to $\gamma(f)$ and setting it to zero. Analogously to Section 4.1.3, we then obtain that an estimate of $\gamma(f)$ based on the compressed observations is

$$\hat{\gamma}(f) = \frac{\mathbf{w}^H(f)\mathbf{B}\mathbf{w}(f)}{\frac{1}{M}\sum_m \mathrm{tr}((\mathbf{\Psi}_m \mathbf{w}(f)\mathbf{w}^H(f)\mathbf{\Psi}_m^H)^2)}. \tag{13}$$

Contrarily to (11), the denominator of the Capon version of $\hat{\gamma}(f)$ is not constant because it depends on the design template $\mathbf{w}(f)$, which at its turn depends on the available data and the analysis frequency f. As X is sparse, when the spectral analysis is focused on sparse bands, the Capon template will be more conservative rather than in areas with signal presence, where the Capon template will relax the second-lobes on sparse areas and provide a more accurate estimate of the actual power. As a result, the Capon spectral estimate is convenient for detectors aimed at finding the spectral edges.

Solving for the optimal template from (13) is not feasible due to the presence of the template in the denominator. Here, we provide a simplification consisting the following asymptotic result. If the dataset is large enough, by making $M \to \infty$, the denominator in (13) obeys the ergodicity property

$$\lim_{M\to\infty} \frac{1}{M} \sum_m \mathrm{tr}((\mathbf{\Psi}_m \mathbf{w}(f)\mathbf{w}^H(f)\mathbf{\Psi}_m^H)^2) \longrightarrow \mathrm{tr}\left((\mathbf{w}(f)\mathbf{w}^H(f)\mathbb{E}[\Psi^H\Psi])^2\right),$$

where Ψ is the $K \times N$ random process from which the compression matrices are drawn. For the pseudo-random pinning matrices, it is shown that $\mathbb{E}[\Psi^H\Psi] = \frac{K}{N}\mathbf{I}$. As a result, as $\|\mathbf{w}(f)\|^2 = N$, the denominator results in $\left(\frac{K}{N}\right)^2 \|\mathbf{w}(f)\|^2 = K^2$, and then (13) becomes

$$\hat{\gamma}(f) = \frac{1}{K^2}\mathbf{w}(f)\mathbf{B}\mathbf{w}^H(f),$$

analogous to the Blackman-Tuckey (15) for $\mathbf{w}(f)$ the steering vector.

Furthermore, the design of the template becomes equivalent to optimize the numerator of the spectral estimate, i.e.,

$$\hat{\mathbf{w}}(f) = \arg\min_{\mathbf{w}(f)} \mathbf{w}^H(f)\mathbf{B}\mathbf{w}(f) \tag{14a}$$

$$\text{s.t.} \quad \mathbf{w}^H(f)\mathbf{s}(f) = N. \tag{14b}$$

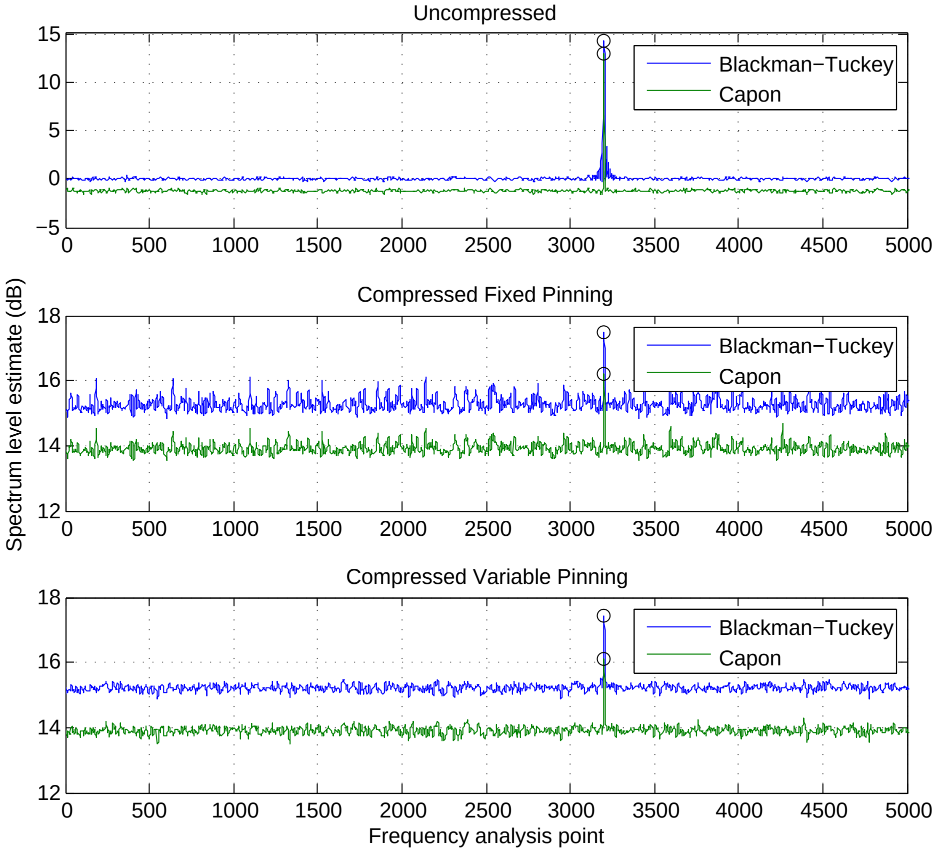

Figure 2: Blackman-Tuckey and Capon power level estimates in a noisy environment with compressed-sampling at $1/16$ of the Nyquist rate ($N = 16K$).

The optimal expression of the Capon filter, i.e., the solution of the former constraint convex optimization problem, is obtained by taking the derivative of the Lagrangian $\mathcal{L}(\mathbf{w}(f)) = \mathbf{w}^H(f)\mathbf{B}\mathbf{w}(f) + \lambda(N - \mathbf{w}^H(f)\mathbf{s}(f))$ and setting it to zero. This gives $\hat{\mathbf{w}}(f) = \lambda\mathbf{B}^\dagger\mathbf{s}(f)$, where the Lagrange multiplier λ is set to satisfy the constraint in (14b), i.e., $\lambda^{-1} = N\mathbf{s}(f)^H\mathbf{B}^\dagger\mathbf{s}(f)$, and $\mathbf{B}^\dagger$ denotes the Moore-Penrose pseudoinverse of the rank-deficient matrix $\mathbf{B}$. Finally, after some mathematical manipulations, we obtain that the Capon spectral estimate for compressed sampling is given by

$$\hat{\gamma}(f) = \frac{N^2}{K^2}\frac{1}{\mathbf{s}(f)^H\mathbf{B}^\dagger\mathbf{s}(f)} \tag{15}$$

It is straightforward to check that under the case of no compression, (15) becomes the traditional Capon filter [89], i.e., $\hat{S}_X(f) = (\mathbf{s}(f)^H\mathbf{R}_X^{-1}\mathbf{s}(f))^{-1}$.

Figure 2 illustrates a toy example on spectral analysis of a complex signal composed by a pure tone at discrete frequency $\omega = 5\pi/4$ with additive Gaussian noise. The comparison between the spectral analysis based on uncompressed or compressed observations with a compression rate of $1/16$ shows the accuracy degradation incurred by the compression. Furthermore, the *noise enhancement* produced by the compression is an important effect to highlight. The reason for such phenomenon is that the white noise before compression is spectrally folded and what is seen at the output of the compression block is another white noise with increased level. To overcome this effect, a denoising process is required, which is discussed in more detail in Section 4.3.

4.2 Wideband Spectrum Sensing

4.2.1 System Description

We consider the spectrum sensing problem with M independent observations of N-size samples, $\mathbf{X} \doteq (\mathbf{x}_1, \ldots, \mathbf{x}_M)$, which consists of only noise samples when the primary user is not transmitting on the sensed channel, denoted by hypothesis $\mathcal{H}_0$, and of noise and signal samples when the channel is used by the primary service, denoted by hypothesis $\mathcal{H}_1$. The spectrum sensing problem is then cast as a multiple hypothesis testing, given by

$$\begin{array}{lcl}\mathcal{H}_0 & : & \mathbf{x}_m = \mathbf{w}_m \\ \mathcal{H}_1 & : & \mathbf{x}_m = \mathbf{s}_m + \mathbf{w}_m,\end{array} \tag{16}$$

for $1 \leq m \leq M$, where $\mathbf{w}$ is the zero-mean complex Gaussian noise with covariance matrix $\mathbf{C}_W \doteq \mathbb{E}[\mathbf{w}\mathbf{w}^H]$, and $\mathbf{s}$ denotes the primary user signal present and the cognitive ratio receiver, which is modeled as a zero-mean complex Gaussian variable with covariance matrix $\mathbf{C}_S \doteq \mathbb{E}[\mathbf{s}\mathbf{s}^H]$. We will also consider the white noise or white signal assumption; in such a case $\mathbf{C}_W = \sigma_W^2\mathbf{I}$ and $\mathbf{C}_S = \sigma_S^2\mathbf{I}$, where σ_W^2 and σ_S^2 are the noise and signal variances, and $\mathbf{I}$ is the identity matrix. In the problem in hand, if the number of observations is sufficiently large, it is a valid assumption that both noise and signal are normally distributed. Moreover, while facilitating the analysis, it is also reasonable because usually there is no line-of-sight (LOS) path between the cognitive radio receiver and the primary user transmitter. As a result, the resulting signal is the superposition of no-LOS signals and approximates Gaussian distribution as the number of observations is sufficienty large according to the central limit theorem [90].

We are interested in detecting the presence of the signal s_n at a given wide band based on the available observations at the cognitive radio receiver. It is known that the GLRT is asymptotically optimal in the Neyman-Pearson criterion, i.e., to maximize the probability of detection, $\mathbb{P}(\mathcal{H}_1|\mathcal{H}_1)$, for a given probability of false alarm level, $\mathbb{P}(\mathcal{H}_1|\mathcal{H}_0) \leq \alpha$, when the number of observations tends to infinity [44]. Recently, the following result establishes the finite-sample optimality of GLRT [91].

Proposition 1. *For an observation $\mathbf{X}$ and unknown set of parameters $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$ with priors $\mathbf{\Theta}_0 \sim \pi(\mathbf{\Theta}_0)$ and $\mathbf{\Theta}_1 \sim \pi(\mathbf{\Theta}_1)$, respectively, the optimal test under the Neyman-Pearson criterion for deciding between hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by the GLRT*

$$L(\mathbf{X}) \doteq \frac{\pi(\hat{\mathbf{\Theta}}_1)p(\mathbf{X}|\mathcal{H}_1, \hat{\mathbf{\Theta}}_1)}{\pi(\hat{\mathbf{\Theta}}_0)p(\mathbf{X}|\mathcal{H}_0, \hat{\mathbf{\Theta}}_0)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma, \tag{17}$$

where the estimates of the unknown parameters are given by

$$\begin{array}{l}\hat{\mathbf{\Theta}}_0 = \arg\max_{\mathbf{\Theta}_0} \pi(\mathbf{\Theta}_0)p(\mathbf{X}|\mathcal{H}_0, \mathbf{\Theta}_0) \\ \hat{\mathbf{\Theta}}_1 = \arg\max_{\mathbf{\Theta}_1} \pi(\mathbf{\Theta}_1)p(\mathbf{X}|\mathcal{H}_1, \mathbf{\Theta}_1),\end{array} \tag{18}$$

and the threshold γ is such that the probability of false alarm satisfies

$$\mathbb{P}(\mathcal{H}_1|\mathcal{H}_0) = \mathbb{P}(L(\mathbf{X}) \geq \gamma|\mathcal{H}_0) = \alpha. \tag{19}$$

We next discuss the derivation of the optimal GLRT detectors in wideband cognitive radio. We define the sample covariance matrix of the observations as the $N \times N$ Hermitian matrix $\mathbf{R}_X \doteq \frac{1}{M}\mathbf{X}\mathbf{X}^H$.

4.2.2 Generalized Likelihood Ratio Tests for Wideband Spectrum Sensing

All Parameters Known. Let the cognitive radio receiver have perfect knowledge on the covariance matrices of the signal and noise, and no prior information on the hypotheses. In this case, the estimates (18) are substituted by their exact value and the test (17) becomes the classical likelihood ratio test (LRT) for Gaussian variables.

Proposition 2. *The optimal test for Gaussian noise and signal with known covariance matrices is the estimator-correlator, given by [44, Equation (5.16)]*

$$T_1(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) = \operatorname{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{R}_X\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_1. \tag{20}$$

Proof. A remark on the proof is given in Appendix A. □

Proposition 2 is a classical detection result, which states that the optimal detector is obtained by correlating the observation data with the output of the Wiener filter or MMSE estimate of the signal, the term $\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{x}_m$, with a whitener term for the noise. Equation (20) can be particularized for white noise, i.e., $\mathbf{C}_W = \sigma_W^2\mathbf{I}$, deriving the equivalent test [44, Equation (5.5)] $T(\mathbf{X}|\mathbf{C}_S, \sigma_W^2) = \operatorname{tr}(\mathbf{C}_S(\mathbf{C}_S + \sigma_W^2\mathbf{I})^{-1}\mathbf{R}_X)$.

Theorem 1. *The optimal LRT for the hypothesis problem* (16) *when asymptotically approaching to low-SNR is given by*

$$T_1(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) \rightarrow \operatorname{tr}(\mathbf{C}_W^{-1}\mathbf{C}_S\mathbf{C}_W^{-1}\mathbf{R}_X) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_1', \tag{21}$$

Proof. The outline of the proof is given in Appendix B. □

The low-SNR regime test (21) also accounts for the correlation between the received data and the signal statistics. However, the test (20) is consistent with the SNR, reflected in by the term $(\mathbf{C}_S + \mathbf{C}_W)^{-1}$, which takes into consideration both signal thermal noise, being the latter predominant at low-SNR, and the signal leakage predominant at high-SNR. A case of interest is when the noise is white, for which Equation (21) is particularized to the test

$$T_1(\mathbf{X}|\mathbf{C}_S, \sigma_W^2) = T_1(\mathbf{X}|\mathbf{C}_S) \rightarrow \operatorname{tr}(\mathbf{C}_S\mathbf{R}_X) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_1''$$

which does not depend on the noise variance. In this case, the detector simply correlates $\mathbf{C}_S$ with the sample correlation matrix of the observations, namely, computes the norm of the projection of the sample covariance matrix over the signal space.

For the tests T_1 derived in this Section with known $\mathbf{C}_S$ and $\mathbf{C}_W$, the family of thresholds γ_1 is set to satisfy the false alarm level α in (19).

Unknown Signal Covariance Matrix, Known Noise Covariance Matrix. The knowledge of $\mathbf{C}_W$ is a general assumption in cognitive radio networks, as a period of no transmission by the primary or secondary users is regularly preset so that each receiver is able to calibrate the noise present at the receiver. In this work, we assume that the cognitive users have knowledge on $\mathbf{C}_W$, and not focus on how this information is obtained. This information aids the test (17) as a prior information to estimate the signal covariance matrix according to (18). A general ML estimate for $\mathbf{C}_S$ is hereby obtained, and simpler detectors are only derived under the white-noise assumption, or for large data records.

Theorem 2. *The optimal test for the cognitive radio detection problem* (16) *when the noise covariance matrix is known and the signal covariance matrix is unknown is given by*

$$T_2(\mathbf{X}|\mathbf{C}_W) = \mathrm{tr}\left(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{R}_X\right) - \ln\det(\hat{\mathbf{C}}_S + \mathbf{C}_W) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_2, \tag{22}$$

where the ML estimate of the signal covariance matrix, when the noise covariance matrix is known by the cognitive radio, is given by $\hat{\mathbf{C}}_S = \mathbf{U}_A\mathbf{\Lambda}_A^+\mathbf{U}_A^H$, *with* $\mathbf{U}_A\mathbf{\Lambda}_A\mathbf{U}_A^H$ *being the eigenvalue decomposition of the correlation matrix* $\mathbf{A} \doteq \mathbf{R}_X - \mathbf{C}_W$, *and the superscript in* $\mathbf{\Lambda}_B^+$ *denoting the diagonal-wise operation* $\max(\lambda_n(\mathbf{A}), 0)$. *Furthermore, the GLRT* (17) *can be simplified for the following two specific cases. On the one hand, for large data records, the optimal test* (22) *becomes*

$$T_2(\mathbf{X}|\mathbf{C}_W) = \sum_{n\in\mathcal{A}_+}\left(\ln\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)} - \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_2' \tag{23}$$

whereas, on the other hand, under the white-noise assumption, the optimal test becomes

$$T_2(\mathbf{X}|\mathbf{C}_W) = \sum_{n\in\mathcal{A}_+}\left(\ln\frac{\lambda_n(\mathbf{R}_X)}{\sigma_W^2} - \frac{\lambda_n(\mathbf{R}_X)}{\sigma_W^2}\right) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_2'', \tag{24}$$

where in (23) *and* (24), *the indexes set* $\mathcal{A}_+$ *is defined as* $\mathcal{A}_+ \doteq \{n \,:\, \lambda_n(\mathbf{R}_X) > \lambda_n(\mathbf{C}_W) > 0\}$.

Proof. The proof of results (22), (23) and (24) is given in Appendix D. □

We note that test (22) is a modified version of the estimator-correlator (20), by placing the ML estimate of $\mathbf{C}_S$ instead of its known value and subtracting the norm $\hat{\mathbf{C}}_S$ onto the inverted noise covariance matrix. On the other hand, the simpler tests (23) and (24) implement an SNR comparison which takes into account solely the part of the received signal that carries energy above the noise level, i.e., those modes associated to eigenvalue larger noise level.

The GLRT with unknown signal covariance matrix can be further simplified for the low-SNR regime.

Theorem 3. *The optimal test for the cognitive radio problem* (16) *when asymptotically approaching to low-SNR when the noise covariance matrix and the signal covariance matrix is unknown is given by*

$$T_2(\mathbf{X}|\mathbf{C}_W) \to \mathrm{tr}(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S\mathbf{C}_W^{-1}\mathbf{R}_X) - \mathrm{tr}(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_2''', \tag{25}$$

where the ML estimate of the signal covariance matrix is given as in Theorem 2. On the other hand, for large data *records the optimal test when approaching to low-SNR becomes*

$$T_2(\mathbf{X}|\mathbf{C}_W) \to -\sum_{n\in\mathcal{A}_+}\left(\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}-1\right)^2 \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_2^*, \tag{26}$$

which is equivalent for the white noise assumption with $\lambda_n(\mathbf{C}_W) = \sigma_W^2$.

Proof. The proof is outlined in Appendix E. □

Note that the test (25) is also a modified versions of the low-SNR estimator-correlator (21), respectively, by placing the ML estimate of $\mathbf{C}_S$ instead of its known value. Similarly, the test given in (26) asymptotically computes the sum of the squared SNRs at each observation mode.

We see from (23), (24) and (26) that the eigenvalue decomposition of the sample and noise covariance matrices is required, which is usually a high-demanding operation. For large data records and Toeplitz stationary correlation matrices, this can be relaxed by computing $\lambda_n(\mathbf{R}_X) = S_X(f_n)$, where $S_X(f)$ is the spectral density of the observations and $f_n \doteq n/N$; and similarly for $\lambda_n(\mathbf{C}_W)$. Contrarily, in the general form of test T_2, namely tests (22) and (25) it is unavoidable the eigenvalue decomposition operation when estimating $\mathbf{C}_S$.

For the tests T_2 derived in this Section with known $\mathbf{C}_W$ and unknown $\mathbf{C}_S$, the family of thresholds γ_2 is set to satisfy the false alarm level α in (19).

Unknown Noise Covariance Matrix, Known Signal Covariance Matrix. We next consider the alternative perspective in which the cognitive radio receiver has knowledge on the signal covariance matrix, i.e., the statistics of the potential primary service that transmits over the sensed band. For a narrow band spectrum sensing, we will assume that $\mathbf{C}_S$ represents a single channel or service. According to the test (17), the noise covariance matrix $\mathbf{C}_W$ is to be ML estimated under both hypothesis, which we denote by $\hat{\mathbf{C}}_{W_0}$ and $\hat{\mathbf{C}}_{W_1}$.

Theorem 4. *The optimal test for the cognitive radio detection problem* (16) *when the signal covariance matrix is known and the noise covariance matrix is unknown is given by*

$$T_3(\mathbf{X}|\mathbf{C}_S) = -\ln\det(\mathbf{C}_S + \hat{\mathbf{C}}_{W_1}) + \ln\det(\hat{\mathbf{C}}_{W_0}) - \mathrm{tr}\left((\mathbf{C}_S + \hat{\mathbf{C}}_{W_1})^{-1}\mathbf{R}_X\right) \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_3, \tag{27}$$

where the ML estimates of the noise covariance matrix under hypotheses $\mathcal{H}_0$ *and* $\mathcal{H}_1$ *are given by* $\hat{\mathbf{C}}_{W_0} = \mathbf{R}_X$*,*$\hat{\mathbf{C}}_{W_1} = \mathbf{U}_B\mathbf{\Lambda}_B^+\mathbf{U}_B^H$*, respectively. Here,* $\mathbf{U}_B\mathbf{\Lambda}_B\mathbf{U}_B^H$ *is the eigenvalue decomposition of the correlation matrix* $\mathbf{B} \doteq \mathbf{R}_X - \mathbf{C}_S$ *and the superscript in* $\mathbf{\Lambda}_B^+$ *denotes the diagonal-wise operation* $\max(\lambda_n(\mathbf{B}), 0)$*. Similarly to Section 4.2.2, the detector can be simplified for large data records, for which the GLRT is then given by*

$$T_3(\mathbf{X}|\mathbf{C}_S) = \sum_{n\in\mathcal{B}_+}\left(\ln\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_S)} - \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_S)}\right) \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_3, \tag{28}$$

the set of indexes $\mathcal{B}_+$ *is defined as* $\mathcal{B}_+ \doteq \{n \,:\, \lambda_n(\mathbf{R}_X) > \lambda_n(\mathbf{C}_S) > 0\}$.

Proof. The proof is outlined in Appendix F. □

Although considering the low-SNR regime, the test (27) does not allow further simplification because two different estimators for the noise variance are provided. However, for large data records, we state the following Theorem.

Theorem 5. *The optimal test for the cognitive radio problem* (16) *when the noise covariance matrix is unknown at the low-SNR regime is givem by*

$$T_3(\mathbf{X}|\mathbf{C}_S) \to -\sum_{n\in\mathcal{B}_+}\left(\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_S)}-1\right)^2 \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_3'. \tag{29}$$

Proof. The proof has been omitted, and is analogous to the proof of Theorem 3. □

Note that the tests derived in Section 4.2.2 and 4.2.2 are very similar, as it compares the information on the structure of the observations to the structure of the known part of the observations, in terms of the eigenvalues. In both cases, the larger the eigenvalues of the sample covariance matrix are relatively to the noise level or to the signal level, the larger the probability of a primary user occupying the sensed band. Here, for large data records, the eigenvalue decomposition calculation can also be relaxed by computing $\lambda_n(\mathbf{R}_X) = S_X(f_m)$, where $S_X(f)$ is the spectral density of the observations and $f_n \doteq n/N$; and similarly for $\lambda_n(\mathbf{C}_S)$.

We now consider the prior knowledge that the noise is white.

Theorem 6. *Under the white noise assumption, the optimal test for the cognitive radio detection problem* (16) *is given by*

$$T_3(\mathbf{X}|\mathbf{C}_S) = M\ln\hat{\sigma}^2_{W_0} - \ln\det(\mathbf{C}_S+\hat{\sigma}^2_{W_1}\mathbf{I}) - \mathrm{tr}\left((\hat{\sigma}^2_{W_1}\mathbf{I}+\mathbf{C}_S)^{-1}\mathbf{R}_X\right) + \hat{\sigma}^{-2}_{W_0}\mathrm{tr}\mathbf{R}_X \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_3'', \tag{30}$$

where the ML estimates of the noise variances under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ for the cognitive radio problem (16) *when the signal covariance matrix is known are given by [44] $\hat{\sigma}^2_{W_0} = \frac{1}{N}\mathrm{tr}(\mathbf{R}_X)$, $\hat{\sigma}^2_{W_1} = \frac{1}{N}\left(\mathrm{tr}(\mathbf{R}_X)-\mathrm{tr}(\mathbf{C}_S)\right)^+$, respectively. A simple detector is not obtained as the sample covariance matrix and the signal covariance matrix could not share eigenvectors. However, a simplification is considered at the low-SNR regime. The optimal test for the cognitive radio problem* (16) *when the white noise covariance is unknown at the low-SNR regime is simplified to*

$$T_3(\mathbf{X}|\mathbf{C}_S) \to \ln\frac{\hat{\sigma}^2_{W_0}}{\hat{\sigma}^2_{W_1}} - \frac{\hat{\sigma}^2_{W_0}}{\hat{\sigma}^2_{W_1}} \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_3'''. \tag{31}$$

Proof. An outline of the proof is given in Appendix G. □

Again, the test (31) is proportional to the sensed SNR incorporating the knowledge on the signal covariance matrix. However, notice that the restriction of white noise extracts less information from $\mathbf{C}_S$, as only the trace operator is involved in the ML estimate $\hat{\sigma}^2_{W_1}$.

For the tests T_3 derived in this Section with known $\mathbf{C}_S$ and unknown $\mathbf{C}_W$, the family of thresholds γ_3 is set to satisfy the false alarm level α in (19).

Unknown Signal and Noise Covariance Matrices. We now discuss the situation in which the cognitive radio receiver needs to estimate both $\mathbf{C}_S$ and $\mathbf{C}_W$. A totally blind detection is not possible when both signals are Gaussian, as it is easily seen that both ML estimates $\hat{\mathbf{C}}_W$ in $\mathcal{H}_0$, and $\hat{\mathbf{C}}_W+\hat{\mathbf{C}}_S$ in $\mathcal{H}_1$

become the same [47]. In other words, some prior information on the statistics or structure of the covariance matrices is to be known by the cognitive radio receivers in order to successfully perform blind detection from the observations $\mathbf{X}$.

The following result estates the GLRT with unknown signal and noise covariance matrices when the noise is white.

Proposition 3. *The optimal test for the cognitive radio problem* (16) *when the colored signal covariance matrix and the white noise variance are unknown is given by [46]*

$$T_4(\mathbf{X}) = \frac{\frac{1}{N}\mathrm{tr}(\mathbf{R}_X)}{\det(\mathbf{R}_X)^{1/N}} \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_4. \tag{32}$$

Proof. The proof is outlined in Appendix H. □

The test (32) can be regarded as the arithmetic-mean to geometric-mean of the eigenvalues of the correlation matrix. We note that this test is a good measure of the spread of the set of eigenvalues. The more unequal the values, the larger the value of the test. For white signal and white noise, both means are equal and (32) is not informative.

We now discuss an additional test which implements (32) in an alternative fashion.

Theorem 7. *The test that computes the dispersion of the eigenvalues of the sample covariance matrix is given by*

$$T_4(\mathbf{X}) \approx \frac{1}{2p^2N}\sum_{n=1}^{N}(\lambda_n(\mathbf{R}_X)-p)^2 \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma_4', \tag{33}$$

where $p \doteq \frac{1}{N}\mathrm{tr}(\mathbf{R}_X)$ *is the estimated average power level.*

Proof. The outline of the proof is given in Appendix J. □

Note that both (32) and (33) are a dispersion measure, which is equal to zero in the case the sample covariance matrix shows no dispersion in terms of eigenvalues, and positive value if there exists dispersion. The advantage of (32) is that the eigenvalue decomposition of $\mathbf{R}_X$ is not required. For stationary processes, though, the computation of $\lambda_n(\mathbf{R}_X)$ for (33) can be done spectrally, i.e., $\lambda_n(\mathbf{R}_X) = S_X(f_n)$, where $S_X(f)$ is the spectral density of the observations and $f_n \doteq n/N$.

For the tests T_4 derived in this Section with unknown $\mathbf{C}_S$ and $\mathbf{C}_W$, the family of thresholds γ_4 is set to satisfy the false alarm level α in (19).

4.2.3 A Correlation-Matching Approach

The method of ML, despite its theoretical appeal, is often difficult to implement [92]. Analytic solutions to the maximization problem are available only in a few circumstances. Numerical solutions are possible in many cases, but may be difficult to program, computationally intensive, or both. We show that the sample covariance matrix of the received observations is a sufficient statistic for the detection problem, and propose the application of the correlation-matching technique as alternative to the ML estimation for the GLRT. The correlation-matching is a least-squares fitting in terms of second-order statistics which behaves as an approximation to ML at the low-SNR regime with asymptotic large number of observations. When the noise level is high, second-order statistics in ML preserve the smallest estimation noise, resulting to an ML function practically insensible to the statistics of higher order in the low-SNR regime [43]. Hence, we will investigate the correlation-matching approach for spectrum sensing as it provides closed-form and low-complex solutions.

We now discuss a correlation-matching detector which relaxes the knowledge on the signal covariance matrix up to a scaling factor. Correlation-matching based techniques gained attention for parameter estimation as an alternative to ML and subspace methods [43].

Theorem 8. *Assume that $\mathbf{C}_S = \sigma_S^2 \mathbf{C}_0$, where $\mathbf{C}_0$ is the normalized signal covariance matrix, diag$(\mathbf{C}_0) = N$, and perfectly known to the cognitive radio. The correlation-matching estimates of the signal and noise variances under $\mathcal{H}_1$ are given by*

$$\hat{\sigma}_S^2 = \left(\frac{\mathrm{tr}((\mathbf{C}_0 - \mathbf{I})\mathbf{R}_X)}{\mathrm{tr}(\mathbf{C}_0^2) - N}\right)^+ \tag{34}$$

$$\hat{\sigma}_{W_1}^2 = \left(\frac{\mathrm{tr}(\mathbf{R}_X)}{N} - \hat{\sigma}_S^2,\right)^+ \tag{35}$$

whereas the correlation-matching estimate of the noise variance under $\mathcal{H}_0$ is equivalent to the ML estimate of Theorem 6. The estimates can be placed to the GLRT (17) *to derive the test*

$$T_3(\mathbf{X}|\mathbf{C}_S) = M \ln \hat{\sigma}_{W_0}^2 - \ln \det(\hat{\sigma}_S^2 \mathbf{C}_0 + \hat{\sigma}_{W_1}^2 \mathbf{I}) - \mathrm{tr}\left((\hat{\sigma}_{W_1}^2 \mathbf{I} + \hat{\sigma}_S^2 \mathbf{C}_0)^{-1} \mathbf{R}_X\right) + \hat{\sigma}_{W_0}^{-2} \mathrm{tr} \mathbf{R}_X \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_3^{**}. \tag{36}$$

Proof. An outline of the proof is given in Appendix I. □

4.2.4 Multi-Frequency Systems

We now consider an orthogonal multi-frequency system with K equally spaced subcarriers ω_k, $1 \leq k \leq K$. The signal in (16) is then the composition of the K frequency signals, i.e., $\mathbf{s}_m = \sum_{k=1}^K \sigma_k \mathbf{s}_{k,m}$, where $\sigma_k \leq 0$ denotes the amplitude of the k-th signal. From the above definition, the signal covariance matrix under the Gaussian assumption reads $\mathbf{C}_S = \sum_{k=1}^K \sigma_k^2 \mathbf{C}_k$, where $\mathbf{C}_k \doteq \mathbb{E}[\mathbf{s}_k \mathbf{s}_k^H]$. In the following, we further assume that the multi-frequency system employs a known modulation format, the same for all the carriers, so that without loss of generality, $\mathbf{C}_k = \mathbf{C}_0 \odot \mathbf{F}_k$, where $\mathbf{C}_0$ is the signal covariance matrix of the baseband modulation format, and $\mathbf{F}_k$ is the correlation version of the Fourier matrix, i.e., $[\mathbf{F}_k]_{m,n} = \mathbf{e}_k \mathbf{e}_k^H$, with $[\mathbf{e}_k]_m = e^{-\jmath \omega_k m}$. Thus, the covariance matrix of the composite signal is given by

$$\mathbf{C}_S = \mathbf{C}_0 \odot \sum_{k=1}^K \sigma_k^2 \mathbf{F}_k. \tag{37}$$

Equation (37) can be particularized for various frequency systems, such as an OFDM-based model with pilot signals available at a known given frequency positions. In such a case, we have that $\mathbf{C}_S$ is as in (37) but with P out of K frequencies are used to transmit known pilots, that is, $\mathbf{C}_S = \mathbf{C}_0 \odot \sum_{k \in \mathcal{K}} \sigma_k^2 \mathbf{F}_k + \mathbf{C}_P \odot \sum_{p \in \mathcal{P}} \sigma_p^2 \mathbf{F}_p$, where the frequency sets $\mathcal{K}$ and $\mathcal{P}$ denote the carrier indexes of data and pilot signals, respectively, and usually $\mathbf{C}_P$ is a rank-1 pilot covariance matrix, whose frequency amplitudes σ_p^2 are fully known. This model can be exploited in real systems where the knowledge on the pilots is available by nature of the primary user service, hence improving the performance of the testing technique without any extra modification on the primary users' parameters.

Multiple-Hypotheses Testing GLRT. The cognitive radio detection problem (16) when the covariance matrix of the signal has the structure (37) consists of determining which of the K subcarriers are occupied by the primary users. We distinguish the two following approaches.

1. Joint Multiple-Hypothesis Testing. We consider the finite-sample optimal GLRT for multiple-hypothesis testing without prior information on the occupancy. That is [91]

$$L(\mathbf{X}) = \frac{\max_l p(\mathbf{X}|\mathcal{H}_{1l}, \hat{\boldsymbol{\sigma}}_l^2)}{p(\mathbf{X}|\mathcal{H}_0)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma,$$

   where l denotes the index for the $2^K - 1$ possible subhypothesis in $\mathcal{H}_1$, and $\boldsymbol{\sigma}_l^2$ represents the subset of $\boldsymbol{\sigma}^2$ that are activated, to be estimated. The complexity of the system increases exponentially with the number of subcarriers, and the ML estimate problem becomes NP-hard.

2. Carrier-Independent Hypotheses Testing. A more efficient detection is considered by performing individual tests at each subcarrier, while assuming that the $K-1$ resting subcarriers are nuisance parameters. Then, we have the following tests

$$L_k(\mathbf{X}) = \frac{p(\mathbf{X}|\mathcal{H}_{1k}, \hat{\boldsymbol{\sigma}}^2)}{p(\mathbf{X}|\mathcal{H}_{0k}, \hat{\boldsymbol{\sigma}}_{\bar{k}}^2)} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_k, \tag{38}$$

   for $1 \le k \le K$, and where $\hat{\boldsymbol{\sigma}}_{\bar{k}}^2$ denotes $\hat{\boldsymbol{\sigma}}_{\bar{k}}^2 \doteq [\sigma_1^2, \ldots, \sigma_{k-1}^2, 0, \sigma_{k+1}^2, \ldots, \sigma_K^2]$. In this case, K thresholds γ_k must be determined to satisfy the false alarm levels, which can be different at each subcarrier.

We see from the former tests that ML estimates of $\boldsymbol{\sigma}^2$ or subsets of it must be determined in order to perform the detection. The following Theorem states a fundamental result for the former multiple-hypotheses testing problems and derives a normal equation for the ML estimate of vector $\boldsymbol{\sigma}^2$ at the low-SNR regime.

Theorem 9. *The ML estimate of $\boldsymbol{\sigma}^2$ at the low-SNR regime is given by the solution to the set of equations*

$$\sum_{l=1}^{K} \sigma_l^2 \beta_{kl} = \delta_k, \tag{39}$$

for $1 \le k \le K$, and where the coefficients are given by

$$\beta_{kl} = 2\mathrm{tr}(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{C}_l\mathbf{C}_W^{-1}\mathbf{R}_X) - \mathrm{tr}(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{C}_l) \tag{40}$$

$$\delta_k = \mathrm{tr}(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{R}_X) - \mathrm{tr}(\mathbf{C}_W^{-1}\mathbf{C}_k). \tag{41}$$

The optimal GLRT for the cognitive radio problem (16) *employing the multi-frequency model* (37) *is then given* *by*

$$
\begin{aligned}
T_k(\mathbf{X}|\mathbf{C}_W, \mathbf{C}_0) \quad = \quad & \ln\det\left(\mathbf{C}_W + \sum_{l\neq k}\hat{\sigma}_l^2\mathbf{C}_l\right) \\
& \mathrm{tr}\left(\left((\mathbf{C}_W + \sum_{l\neq k}\hat{\sigma}_l^2\mathbf{C}_l\right)^{-1}\mathbf{R}_X\right) \\
& -\ln\det\left(\mathbf{C}_W + \sum_{k=1}^{K}\hat{\sigma}_k^2\mathbf{C}_k\right) \\
& -\mathrm{tr}\left(\left((\mathbf{C}_W + \sum_{k=1}^{K}\hat{\sigma}_k^2\mathbf{C}_k\right)^{-1}\mathbf{R}_X\right) \\
& \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1}\gamma_k
\end{aligned}
\tag{42}
$$

where the set $\hat{\sigma}_{l\neq k}^2$ is the solution to $(K-1)\times(K-1)$ system (39) *by setting $\sigma_k^2 = 0$, whereas the estimates $\hat{\sigma}_k^2$ are the solution to the $K\times K$ system* (39)*. The set of thresholds γ_k are chosen to satisfy the false alarm level α, which we assume is fixed for all the carriers.*

Proof. See Appendix K. □

In the appendix, we show that the exact MLE for the power levels σ_k^2 are given by the solution of the nonlinear equation

$$
\mathrm{tr}\left(\left(\mathbf{C}_W + \sum_{l=1}^{K}\sigma_l^2\mathbf{C}_l\right)^{-1}\mathbf{C}_k\right) = \mathrm{tr}\left(\left(\mathbf{C}_W + \sum_{l=1}^{K}\sigma_l^2\mathbf{C}_l\right)^{-1}\mathbf{C}_k\left(\mathbf{C}_W + \sum_{l=1}^{K}\sigma_l^2\mathbf{C}_l\right)^{-1}\mathbf{R}_X\right), \tag{43}
$$

for $1\leq k\leq K$. We first observe that Equation (43) forces the ML estimate of $\boldsymbol{\sigma}^2$ to be the set of frequency powers that better shape the signal covariance matrix and the sample covariance matrix across all the frequencies. Let $\mathbf{C}_X = \hat{\mathbf{C}}_S + \mathbf{C}_W$ be the observations covariance matrix where the estimate of the signal covariance matrix is defined from the estimate of the frequency powers as $\hat{\mathbf{C}}_S = \sum_{l=1}^{K}\hat{\sigma}_l^2\mathbf{C}_l$. Then, Equation (43) can be read as $\mathrm{tr}\left(\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H\right) = \frac{1}{N}\mathrm{tr}\left(\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H\tilde{\mathbf{X}}\tilde{\mathbf{X}}^H\right)$, where $\tilde{\mathbf{x}}_k \doteq \mathbf{C}_X^{1/2}\mathbf{x}_k$, and $\tilde{\mathbf{X}} \doteq \mathbf{C}_X^{1/2}\mathbf{X}$. In other words, the ML estimate of $\boldsymbol{\sigma}^2$ performs an optimal signal-level fitting of $\mathbf{C}_X$ to $\mathbf{R}_X$ by whitening the data with $\mathbf{C}_X^{1/2}$.

In the low-SNR regime, we further see that β_{kl} computes the whitened cross-correlation between the data format at the k-th and l-th subcarriers with the sample covariance matrix, whereas δ_k determines the part of the k-th subcarrier signal present in the observations. Equation (39) can be efficiently solved in a matrix manner exploiting the properties of the coefficients β_{kl} and δ_k given the model (37). That is $\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{\beta}^{-1}\boldsymbol{\delta}$, where $[\boldsymbol{\beta}]_{kl} = \beta_{kl}$ and $[\boldsymbol{\delta}]_k = \delta_k$.

Asymptotic Performance. We now discuss the performance of the ML estimator (39) in the asymptotic case with $N\to\infty$, i.e., for large data records at the low-SNR regime. Since the ML estimate of the signal frequency powers is the implicit solution to a system of equations, the computation of the bias is not treatable. Alternatively, the following statement establishes an asymptotic result.

Proposition 4. *The asymptotic set of equations* (39) *for* $N \to \infty$ *are consistent at the low-SNR regime when* $\hat{\boldsymbol{\sigma}}^2 \to \boldsymbol{\sigma}^2$.

Proof. The full discussion is given in Appendix M. □

It is well known that the carrier-independent GLRT tests $L_k(\mathbf{X})$ are asymptotically distributed according to $L_k(\mathbf{X}) \sim \mathcal{X}_r^2$, where r denotes the size of the unknowns. That is, as the other $K-1$ subcarriers are regarded as nuisance parameters, $r = 1$. Hence, the K thresholds γ_k are determined to satisfy the false alarm level $\mathbb{P}(\mathcal{H}_{1k}|\mathcal{H}_{0k}) = \mathbb{P}(L_k(\mathbf{X}) \geq \gamma_k|\mathcal{H}_{0k}) = \alpha_k$. For the multiple-hypothesis testing, the same result applies with $r = K$, i.e., $L(\mathbf{X}) \sim \mathcal{X}_K^2$. We note that the asymptotic statistics of the GLRT are independent on the observation size M and the SNR regime.

Correlation Matching with Known Noise Covariance Matrix. We see that in order to analytically solve the ML estimate of $\boldsymbol{\sigma}^2$, a low-SNR approximation is required. The optimality of ML, hence, will only hold at the low-SNR regime.

On the other hand, we now consider the correlation matching technique. From (39) we observe that second-order statistics are involved in the computation of the discrete spectrum in terms of the sample correlation matrix, i.e., $\mathbf{R}_X$.

Theorem 10. *The solution to the correlation matching problem is given by the following normal equation:*

$$\sum_{l=1}^{K} \sigma_l^2 \tilde{\beta}_{kl} = \tilde{\delta}_k, \tag{44}$$

for $1 \leq k \leq K$*, and where the coefficients are given by*

$$\tilde{\beta}_{kl} = \mathrm{tr}(\mathbf{C}_k \mathbf{C}_l) \tag{45}$$

$$\tilde{\delta}_k = \mathrm{tr}(\mathbf{C}_k \mathbf{R}_X) - \mathrm{tr}(\mathbf{C}_W \mathbf{C}_k). \tag{46}$$

Proof. The proof is given in Appendix L. □

We note that correlation matching and optimal ML estimate for GLRT involve solving very similar equations. In the second case, Equations (45)–(46) can be simplified for white noise, to the following expressions $\tilde{\beta}_{kl} = M\|c_0[0]\|^2 + 2\sum_{n=1}^{N-1}(N-n)\cos\left((\omega_i - \omega_k)n\right)\|c_0[n]\|^2$, and $\tilde{\delta}_k = \mathrm{tr}(\mathbf{C}_k\mathbf{R}_X) - \sigma_W^2 N c_0[0]$.

The MLE provided in (40)–(41) are equivalent to those in (45)–(46) under the white noise assumption, i.e. $\mathbf{C}_W = \sigma_W^2 \mathbf{I}$ and further considering that at the low-SNR regime $\mathbf{R}_X/\sigma_W^2 \to \mathbf{I}$. By doing so, we see that $\beta_{kl} = \sigma_W^4 \tilde{\beta}_{kl}$ and $\delta_{kl} = \sigma_W^4 \tilde{\beta}_{kl}$, becoming equivalent. The main advantage of correlation-matching is that matrix $\tilde{\boldsymbol{\beta}}$ can be computed offline, because it is independent on the received observations. Therefore, solving $\hat{\boldsymbol{\sigma}} = \tilde{\boldsymbol{\beta}}^{-1}\tilde{\boldsymbol{\delta}}$ reduces to the MLE

$$\hat{\sigma}_k^2 = \tilde{\boldsymbol{\beta}}_k^{-1}\tilde{\boldsymbol{\delta}}, \tag{47}$$

where $\tilde{\boldsymbol{\beta}}_k^{-1}$ denotes the k-th row of the inverse of $\tilde{\boldsymbol{\beta}}$. The structure of the signal model can be further exploited according to (37) and compress the ML function to derive the following GLRT

$$\begin{aligned} T_k(\mathbf{X}|\mathbf{C}_0, \sigma_W^2) \quad = \quad & + \ln\det(\sigma_W^2 \mathbf{I} + \sum_{l \neq k} \hat{\sigma}_l^2 \mathbf{C}_l) + \mathrm{tr}\left[(\sigma_W^2 \mathbf{I} + \sum_{l \neq k} \hat{\sigma}_l^2 \mathbf{C}_l)^{-1} \mathbf{R}_X \right] \\ & - \ln\det(\sigma_W^2 \mathbf{I} + \sum_{k=1}^{K} \hat{\sigma}_k^2 \mathbf{C}_k) - \mathrm{tr}\left[(\sigma_W^2 \mathbf{I} + \sum_{k=1}^{K} \hat{\sigma}_k^2 \mathbf{C}_k)^{-1} \mathbf{R}_X \right] \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma_k', \end{aligned} \tag{48}$$

where the set $\hat{\sigma}_{l \neq k}^2$ is the solution to the $(K-1) \times (K-1)$ system (44) by setting $\sigma_k^2 = 0$, and the set $\hat{\sigma}^2$ is the solution to the complete system (44). In (48), the thresholds γ_k are chosen to satisfy the false alarm level α.

Correlation-Matching with Unknown White Noise Variance. We first consider the correlation-matching technique under the white noise assumption with unknown noise variance σ_W^2 for the detection of the multi-frequency signal given in (37). The solution to the correlation-matching problem when sensing the k-th frequency results in

$$\begin{aligned} \mathcal{H}_{0k} \; : \; \hat{\sigma}_{W_0}^2, \hat{\boldsymbol{\sigma}}_{\bar{k}}^2 \quad &= \quad \arg\min \|\sigma_W^2 \mathbf{I} + \mathbf{C}_{\bar{k}} - \mathbf{R}_X\|^2 \\ \mathcal{H}_{1k} \; : \; \hat{\sigma}_{W_1}^2, \hat{\boldsymbol{\sigma}}^2 \quad &= \quad \arg\min \|\sigma_W^2 \mathbf{I} + \mathbf{C}_S - \mathbf{R}_X\|^2, \end{aligned}$$

where $\mathbf{C}_{\bar{k}} \doteq \sum_{l \neq k} \sigma_l^2 \mathbf{C}_l$. By taking the derivative with respect all the variables under both hypotheses, we obtain the following correlation-matching estimates for the noise variances:

$$\hat{\sigma}_{W_0}^2 \quad = \quad \frac{\mathrm{tr}(\mathbf{R}_X)}{M} - \sum_{l \neq k} \hat{\sigma}_l^2 \tag{49a}$$

$$\hat{\sigma}_{W_1}^2 \quad = \quad \frac{\mathrm{tr}(\mathbf{R}_X)}{M} - \sum_{l=1}^{K} \hat{\sigma}_l^2, \tag{49b}$$

where the estimates of the frequency powers is given by the system of equations

$$\mathcal{H}_{0k} \quad : \quad \sum_{l \neq k} \sigma_l^2 \left[\mathrm{tr}(\mathbf{C}_k \mathbf{C}_l) - N \right] = \mathrm{tr}((\mathbf{C}_k - \mathbf{I}) \mathbf{R}_X) \tag{50a}$$

$$\mathcal{H}_{1k} \quad : \quad \sum_{l=1}^{K} \sigma_l^2 \left[\mathrm{tr}(\mathbf{C}_k \mathbf{C}_l) - N \right] = \mathrm{tr}((\mathbf{C}_k - \mathbf{I}) \mathbf{R}_X), \tag{50b}$$

which clearly are a generalization to the narrowband detector. In this particular case, the solution can be written in the form (47) as the coefficients of the unknowns are independent of the observations, namely the elements of the inverse matrix $\boldsymbol{\beta}^{-1}$ can be computed offline.

In conclusion, the optimal test employing the correlation-matching estimates is given by

$$\begin{aligned} T_k(\mathbf{X}|\mathbf{C}_0) &= +\ln\det(\hat{\sigma}^2_{W_0}\mathbf{I}+\sum_{l\neq k}\hat{\sigma}^2_l\mathbf{C}_l)+\mathrm{tr}\left[(\hat{\sigma}^2_{W_0}\mathbf{I}+\sum_{l\neq k}\hat{\sigma}^2_l\mathbf{C}_l)^{-1}\mathbf{R}_X\right] \\ &\quad -\ln\det(\hat{\sigma}^2_{W_1}\mathbf{I}+\sum_{k=1}^{K}\hat{\sigma}^2_k\mathbf{C}_k)-\mathrm{tr}\left[(\hat{\sigma}^2_{W_1}\mathbf{I}+\sum_{k=1}^{K}\hat{\sigma}^2_k\mathbf{C}_k)^{-1}\mathbf{R}_X\right] \gtrless^{\mathcal{H}_1}_{\mathcal{H}_0} \gamma''_k, \end{aligned} \tag{51}$$

where the noise variances under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are given by (49a) and (49b), respectively; the set $\hat{\sigma}^2_{l\neq k}$ is given by the solution of (50a), and $\hat{\sigma}^2_k$, $1\leq k\leq K$, (50b). The K thresholds γ_k in $T_k(\mathbf{X}|\mathbf{C}_0)$ are chosen to satisfy the false alarm level α.

Wald and Rao Tests. The GLRT derived in Sections 4.2.4, 4.2.4, and 4.2.4 involve complex expressions due to the nuisance parameters, and the power level estimates involve solving a set of equations of high dimensionality. The Wald and Rao test [44] are asymptotically equivalent to the GLRT and, in practice, may be easier to compute. They both involve the Fisher information matrix, whose entries at the low-SNR regime are given by

$$[\boldsymbol{\mathcal{I}}(\boldsymbol{\sigma}^2)]_{kl} = \frac{\partial^2 \ln p(\mathbf{X}|\boldsymbol{\sigma}^2)}{\partial\sigma^2_k\partial\sigma^2_l} = \beta_{kl}. \tag{52}$$

The former expression allows simple computation and interpretation of the information matrix. From Equation (40) we observe that those frequency channels with large cross-correlation will have smaller Cramér-Rao bound (CRB). We also note that the information matrix is independent on $\boldsymbol{\sigma}^2$. The following Theorem states the Wald and Rao tests in cognitive radio detection schemes that modeled with the signal covariance structure given by (37).

Theorem 11. *The Wald and Rao tests for the cognitive radio testing problem* (16) *with multi-frequency signal structure* (37) *at the low-SNR regime are given by*

$$T^W_k(\mathbf{X}|\mathbf{C}_W,\mathbf{C}_0) = \frac{\left(\hat{\sigma}^2_k\right)^2}{\left[\boldsymbol{\mathcal{I}}^{-1}\right]_{kk}}, \tag{53}$$

$$T^R_k(\mathbf{X}|\mathbf{C}_W,\mathbf{C}_0) = \left(\sum_{l=1}^{K}\hat{\sigma}^2_{\bar{k}l}\beta_{kl}-\delta_k\right)^2\left[\boldsymbol{\mathcal{I}}^{-1}\right]_{kk}, \tag{54}$$

respectively. In (53), $\hat{\boldsymbol{\sigma}}^2$ *is the ML estimate of* $\boldsymbol{\sigma}^2$*, i.e., the solution to* (39)*, and in* (54) $\boldsymbol{\sigma}^2_{\bar{k}}$ *is defined as in Section 4.2.4, i.e., the ML estimate of the nuisance parameters under* $\mathcal{H}_0$ *by setting* $\sigma^2_k = 0$.

Proof. The proof is given in Appendix N. □

It can be observed that both the Wald and Rao tests compute statistics of order fourth on the signal level. The Wald test simply evaluates the squared value of the estimated power level at the k-th band, weighted by the k-th element in the diagonal of the inverse of the Fisher information matrix, i.e., by the CRB of the estimate. On the other hand, the Rao test evaluates the mismatch between hypothesis $\mathcal{H}_{1k}$ and $\mathcal{H}_{0k}$ in estimating the power at the independent $K-1$ frequencies distinct from the k-th band. In

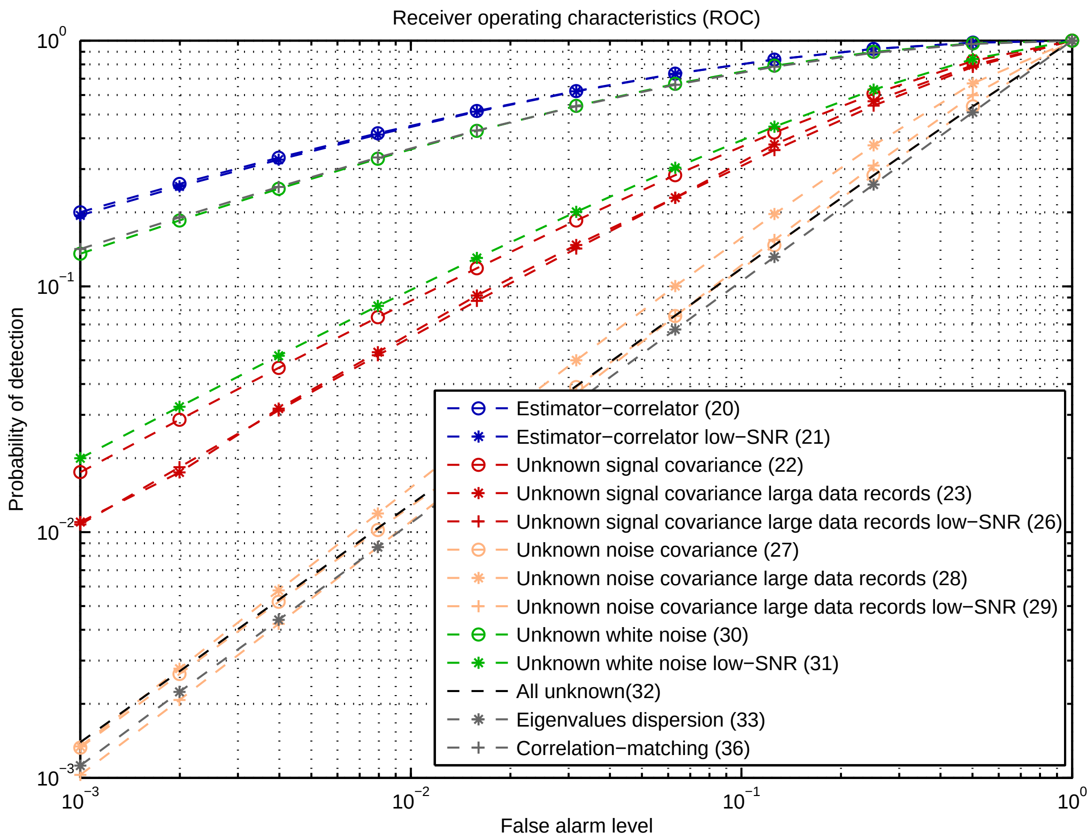

Figure 3: Receiver operating characteristics at the low-SNR regime (SNR = -15 dB), with $M = 32$ and $S = 8$.

other words, $\boldsymbol{\sigma}_{\bar{k}}^2$ is the solution to the $(K-1) \times (K-1)$ system of equations resulting from removing the k-th equation in (39), hence the Rao tests computes the portion of signal that the ML estimate puts into σ_k^2 in $\mathcal{H}_{1k}$ (...). Furthermore, the test also weights the comparison by the inverse Fisher information matrix in the inversely proportional fashion than the Wald test.

4.2.5 Simulation Results

This section provides up-to-date simulation results of the algorithms discussed throughout the paper for the wideband regime. The SNR of detection is defined as the ratio SNR $\doteq \sigma_S^2/\sigma_W^2$, following the normalization presented in the low-SNR approximation given in Appendix C for $\mathbf{C}_S$ and $\mathbf{C}_W$.

Figure 3 depicts the receiver operating characteristics (ROC) in terms of probability of detection versus probability of false in the low-SNR regime. For SNR = -15 dB, we can identify three groups of detectors with similar performance. The first group of high-performance detectors is composed by the estimator-correlator tests, unknown white noise and unknown white noise and signal variance. This justifies the fact that a correct knowledge on the statistics of the cognitive radio detection problem (16) has an important benefit in terms of gain. The statistics involved in this tests are the exact covariance matrices (estimator-correlator), exact signal covariance and white noise assumption, and normalized exact signal covariance with white noise assumption. Interestingly, the detectors with exact signal co-

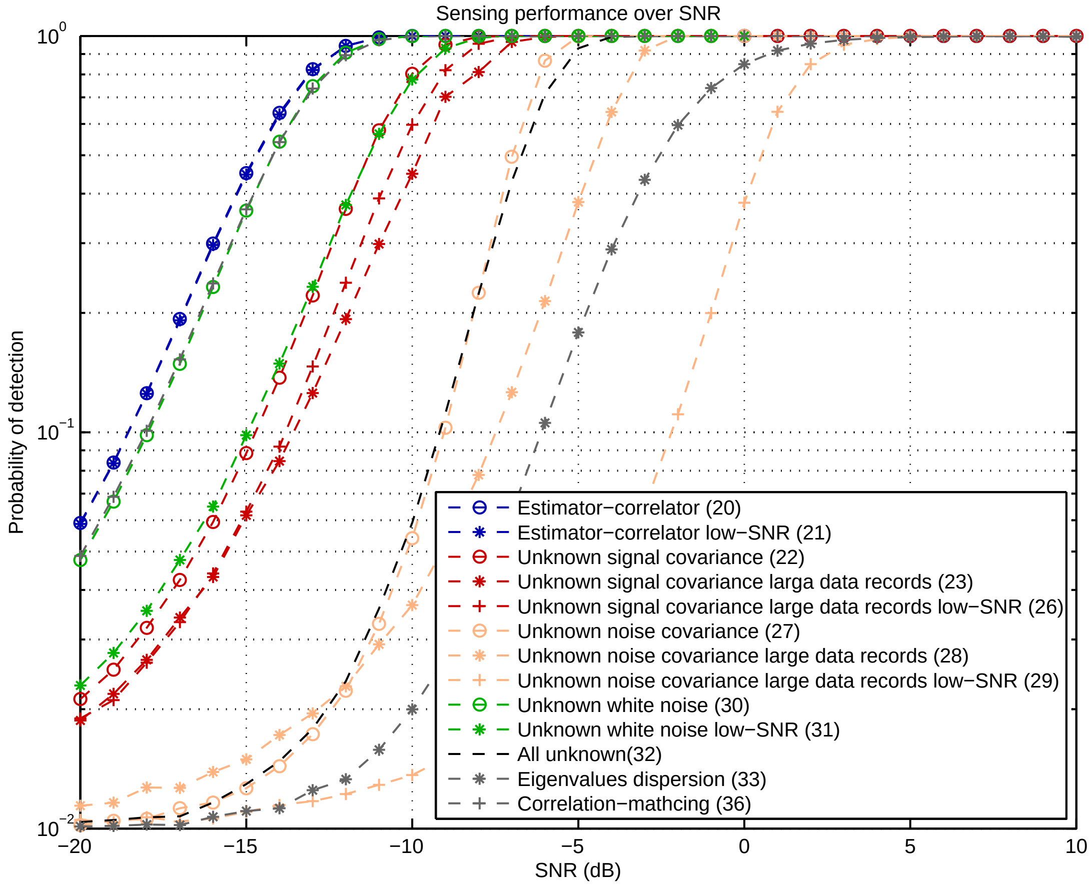

Figure 4: Probability of detection versus average SNR for false alarm level $\alpha = 0.1$, $M = 32$ and $S = 8$.

variance matrix but unknown noise covariance matrix without white noise assumption are located in a ROC performance comparable to the test with all statistics unknown. This is because the information of the noise contained in the observations is white, and this detectors try to erroneously discover some structure on $\mathbf{C}_W$. This last result gives strength to the white noise assumption, as any colored noise can be potentially considered as a primary user signal. The last group of detectors, with performance in between the two above mentioned groups, consists of the detectors with unknown signal covariance, and the low-SNR approximation of the exact signal covariance with unknown white noise. The main conclusion is that the knowledge of the normalized $\mathbf{C}_S$, i.e., $\mathbf{C}_0$, together with the white noise assumption is the optimal set of known statistics for the wideband cognitive radio detection problem (16).

The performance of the detectors versus SNR is depicted in Figure 4 in terms of probability of detection for an observation size of $M = 32$, signal covariance matrix rank $S = 8$, and false alarm level $\alpha = 0.1$. As it can be observed, the performance of the tests is according to the ROC depicted in Figure 3.

Finally, the trade-off between the rank of the covariance matrix and the size of the observations is simulated in Figure 5. It shows the probability of detection at SNR = -15 dB versus the S/M ratio for a false alarm level $\alpha = 0.1$. We can derive some conclusions from the Figure. Firstly, we can identify a set of detectors which are insensitive to the relative rank of the signal covariance matrix. These are,

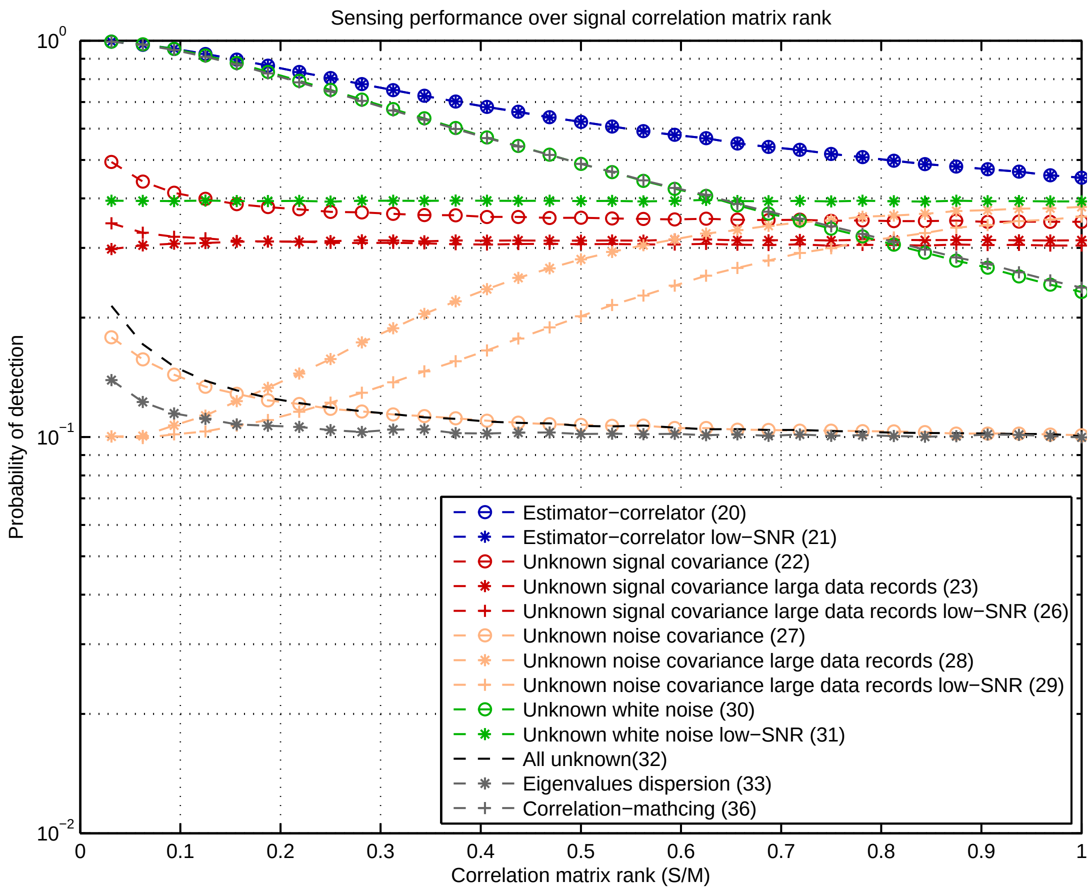

Figure 5: Probability of detection versus S/M ratio at the low-SNR regime (SNR = -15 dB), for false alarm level $\alpha = 0.1$.

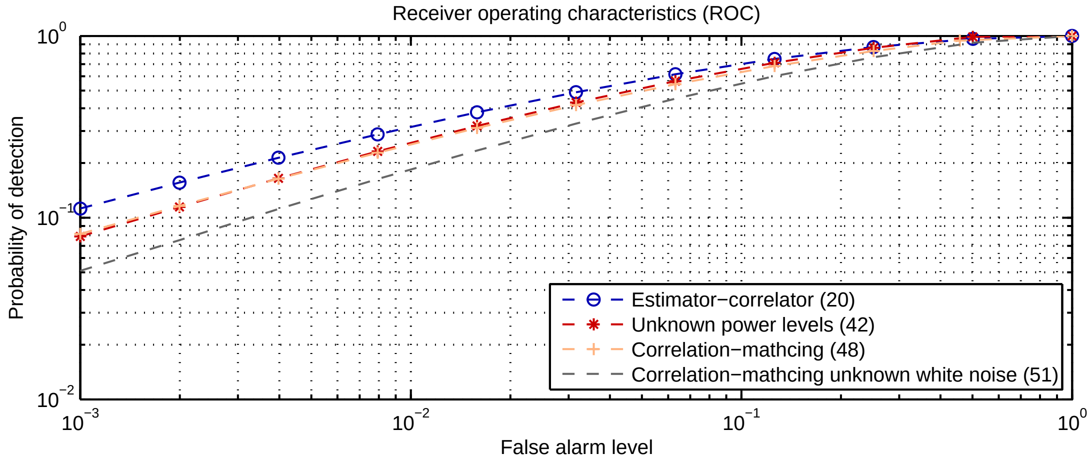

Figure 6: Average receiver operating characteristics at the low-SNR regime (SNR = -15 dB), with $M = 32$, $S = 8$, in a $K = 8$ multi-frequency system.

as expected, the tests with unknown signal covariance —and unknown S—, as well as the test with unknown white noise variance at the low-SNR regime as it simply computes the power of the detected signal. The majority of the detectors show a performance penalty when increasing S/M, because a rank-deficient signal relaxes the detection problem when the noise is white, i.e., full-rank. Notably, the detection of pilot tons, i.e. rank 1, provides the best performance scheme. The two remaining tests, which are the detectors that compares the information of the structure of the covariance matrix in terms of eigenvalues show a performance gain for $S/M \to 1$, as far as the distribution of the eigenvalues of $\mathbf{C}_S$ is enough heterogeneously.

We also provide simulation results on the algorithms derived in Section 4.2.4 for multi-frequency systems. Specefically, we compare the receiver operating characteristics (ROC) curves for the estimator-correlator detector (20), the GLRT (42), and the correlation-matching approaches (48) and (51). In Figure 6 the probability of detection of an arbitrary frequency is evaluated versus the false alarm level α, for a system with $K = 8$ frequencies, observation size of $M = 32$, and $S = \text{rank}(\mathbf{C}_0) = 8$. The performance of the optimal GLRT (42) and the correlation-matching (48) are comparable for the wide range of ROC simulated in the low-SNR regime. We conclude that the knowledge on the noise variance in a wideband system plays a role in the sensing algorithm, and hence the correlation-matching with unknown noise variance (51) incurs performance loss.

4.3 Compressed Correlation-Matching

4.3.1 Power Level-Based Correlation-Matching

In this Section, we propose a correlation-matching approach for detecting the signal power level of the primary user services. Due the compression process and the thermal noise present at the uncompressed observations, the phenomenon of noise enhancement is observed at the output of the reconstruction, as reported in Figure 2. For this reason, we include a denoising process in the formulation in order to diminish the effect of noise enhancement.

Particularly, we are interested in estimating the signal and noise power levels in a correlation-matching approach based on the compressed observations, $\mathbf{y}_m$. On the one hand, we let the signal covariance matrix to be expressed as $\mathbf{R}_X \doteq \gamma\mathbf{R}$, where γ denotes the signal power level associated to the normalized correlation matrix $\mathbf{R}$, with diag$(\mathbf{R}) = N$. On the other hand, let σ^2 be the noise power level present before compression, i.e., $\mathbf{R}_Z = \sigma^2\mathbf{\Sigma}$, with diag$(\mathbf{\Sigma}) = N$. The compressed correlation-matching technique for spectrum sensing is then given by the optimization problem

$$\hat{\gamma}, \hat{\sigma}^2 = \arg\min_{\gamma,\sigma^2} \sum_{m=1}^{M} \left\| \mathbf{y}_m \mathbf{y}_m^H - \mathbf{\Psi}_m(\gamma\mathbf{R} + \sigma^2\mathbf{\Sigma})\mathbf{\Psi}_m^H \right\|^2 . \tag{55}$$

The problem (55) is convex on both γ and σ^2, and the solution is obtained by taking the derivative with respect to the signal and noise power levels and setting it to zero. [The positivity on the variables should be included at first, but here we omit the constraint for the problem in hand and apply the $()^+$ operator at the end.] By taking the derivative with respect to the signal power level γ we obtain

$$\sum_{m=1}^{M} \mathrm{tr}(\mathbf{y}_m\mathbf{y}_m^H\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H) - \gamma\mathrm{tr}((\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H)^2) - \sigma^2\mathrm{tr}(\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H) = 0.$$

Similarly, after taking the derivative with respect to the noise power level σ^2, we obtain

$$\sum_{m=1}^{M} \mathrm{tr}(\mathbf{y}_m\mathbf{y}_m^H\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H) - \sigma^2\mathrm{tr}((\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H)^2) - \gamma\mathrm{tr}(\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H) = 0.$$

Clearly, both equations are coupled and have very similar structure. Hence, the detection on γ and σ^2 is based on the projection of the local compressed observations onto the normalized correlation matrices $\mathbf{R}$ and $\mathbf{\Sigma}$, respectively. After some mathematical manipulations, it can be shown that the optimal signal and noise levels are given by

$$\hat{\gamma} = \frac{\mathrm{tr}(\mathbf{B}\mathbf{R})\mathrm{tr}(\overline{\mathbf{\Sigma}^2}) - \mathrm{tr}(\mathbf{B}\mathbf{\Sigma})\mathrm{tr}(\overline{\mathbf{\Sigma}\mathbf{R}})}{\mathrm{tr}(\overline{\mathbf{R}^2})\mathrm{tr}(\overline{\mathbf{\Sigma}^2}) - \mathrm{tr}^2(\overline{\mathbf{R}\mathbf{\Sigma}})} \tag{56a}$$

$$\hat{\sigma}^2 = \frac{\mathrm{tr}(\mathbf{B}\mathbf{\Sigma})\mathrm{tr}(\overline{\mathbf{R}^2}) - \mathrm{tr}(\mathbf{B}\mathbf{R})\mathrm{tr}(\overline{\mathbf{R}\mathbf{\Sigma}})}{\mathrm{tr}(\overline{\mathbf{\Sigma}^2})\mathrm{tr}(\overline{\mathbf{R}^2}) - \mathrm{tr}^2(\overline{\mathbf{\Sigma}\mathbf{R}})}, \tag{56b}$$

where we have defined the following matrices

$$\mathbf{B} \doteq \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m^H\mathbf{y}_m\mathbf{y}_m^H\mathbf{\Psi}_m \tag{57a}$$

$$\overline{\mathbf{R}} \doteq \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H \tag{57b}$$

$$\overline{\mathbf{\Sigma}} \doteq \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H \tag{57c}$$

$$\overline{\mathbf{R}^2} \doteq \frac{1}{M}\sum_{m=1}^{M}(\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H)^2 \tag{57d}$$

$$\overline{\mathbf{\Sigma}^2} \doteq \frac{1}{M}\sum_{m=1}^{M}(\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H)^2 \tag{57e}$$

$$\overline{\mathbf{R}\mathbf{\Sigma}} \doteq \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H \tag{57f}$$

$$\overline{\mathbf{\Sigma}\mathbf{R}} \doteq \frac{1}{M}\sum_{m=1}^{M}\mathbf{\Psi}_m\mathbf{\Sigma}\mathbf{\Psi}_m^H\mathbf{\Psi}_m\mathbf{R}\mathbf{\Psi}_m^H. \tag{57g}$$

The physical interpretation of (56) can be explained as follows. On the one hand, after dividing the numerator and the denominator of $\hat{\gamma}$ by $\mathrm{tr}(\overline{\mathbf{\Sigma}^2}$, we can see that the detector of the signal power level is of the form $\hat{\gamma} = (\mathrm{tr}(\mathbf{B}\mathbf{R}) - C_1)/C_2$, where the term $\mathrm{tr}(\mathbf{B}\mathbf{R})$ computes the projection of the reconstructed correlation matrix $\mathbf{B}$ onto the normalized signal correlation matrix. To this term, the detector subtracts the term C_1, which accounts for the part of the reconstructed observations that are considered as noise, with a weighting factor that depends on the cross-correlation between $\mathbf{R}$ and $\mathbf{\Sigma}$. The same interpretation can be followed for $\hat{\sigma}^2$ but exchanging the roles of $\mathbf{R}$ and $\mathbf{\Sigma}$.

On the other hand, the common denominator C_2 in (56) is accounting for distinctness of the second-order statistics of the signal and the noise in terms of the Schwartz inequality, since $\mathrm{tr}(\overline{\mathbf{\Sigma}^2})\mathrm{tr}(\overline{\mathbf{R}^2}) \geq \mathrm{tr}^2(\overline{\mathbf{\Sigma}\mathbf{R}})$, with equality when the statistics of the noise and the signal are orthogonal. In such a case, the estimates in (56) become inaccurate because it becomes difficult to distinguish between the noise and the signal contribution in the compressed observations. Conversely, the more distinct are $\overline{\mathbf{R}}$ and $\overline{\mathbf{\Sigma}}$.

A modest simplification occurs under the white noise assumption. The white noise assumption is a robust principle in cognitive radio detection, as any colored spectrum is potentially regarded as primary user activity. [...] In such a case, we particularize Equation (56) with $\mathbf{\Sigma} = \mathbf{I}$, which leads to the following optimal estimates

$$\hat{\gamma} = \frac{\mathrm{tr}(\mathbf{B}(\mathbf{R} - \mathbf{I}))}{\mathrm{tr}(\overline{\mathbf{R}^2}) - K} \tag{58a}$$

$$\hat{\sigma}^2 = \frac{\mathrm{tr}(\mathbf{B})\mathrm{tr}(\overline{\mathbf{R}^2}) - \mathrm{tr}(\mathbf{B}\mathbf{R})\mathrm{tr}(\overline{\mathbf{R}})}{\mathrm{tr}(\overline{\mathbf{R}^2}) - K}. \tag{58b}$$

Now, we can observe the effect of diagonal off-loading. The detector (58a) is of the type $\mathrm{tr}(\mathbf{B}(\mathbf{R} - \mathbf{I})) > \lambda'$, which it can be written alternatively as $\mathrm{tr}(\mathbf{B}(\mathbf{R}_X - N^{-1}\mathrm{tr}(\mathbf{R}_X)\mathbf{I}))$. By defining the *diagonal off-loaded* *correlation matrix* as $\mathbf{R}_X^{\mathrm{OFF}} \doteq \mathbf{R}_X - \mathrm{tr}(\mathbf{R}_X)\mathbf{I}$. The detector can be now interpreted as follows:

$$\sum_{m=1}^{M} \mathbf{y}_m^H \mathbf{\Psi}_m \mathbf{R}_X^{\mathrm{OFF}} \mathbf{\Psi}_m^H \mathbf{y}_m.$$

If $X(t)$ is a stationary process, the correlation-matching approach is an energy detector that takes into account only the part that is not affected by the noise, i.e., the presence of non-zero correlation lags. Conversely, if $X(t)$ is cyclo-stationary, the main diagonal of $\mathbf{R}_X$ is not uniform and the detector evaluates its variability around the mean. Therefore, the detector is a lineal combination between the energetic variability and the correlation in the non-zero lags.

4.3.2 Multi-Frequency Correlation-Matching

We now extend the former formulation for a multi-frequency model in which the primary user signal is a lineal superposition of frequency-orthogonal signals. The signal correlation matrix is then of the type $\mathbf{R}_X = \sum_{l=1}^{L} \gamma_l \mathbf{R}_l$, where each $\mathbf{R}_l$ stands for the normalized correlation matrix at the l-th frequency (channel).

The correlation-matching problem is then formulated as follows. We are interested in detecting the presence of signal from the primary users at each frequency. Let $\boldsymbol{\gamma} \doteq (\gamma_1, \ldots, \gamma_L)$, and σ^2 denote the the noise level prior to compression. Then, the correlation-matching problem is given by

$$\hat{\boldsymbol{\gamma}}, \hat{\sigma}^2 = \arg\min_{\boldsymbol{\gamma}, \sigma^2} \sum_{m=1}^{M} \left\| \mathbf{y}_m \mathbf{y}_m^H - \mathbf{\Psi}_m \left(\sum_{l=1}^{L} \gamma_l \mathbf{R}_l + \sigma^2 \mathbf{\Sigma} \right) \mathbf{\Psi}_m^H \right\|^2.$$

The former problem is also convex on the optimization variables $\gamma_1, \ldots, \gamma_L$ and σ^2. After taking the derivative with respect to γ_i, we obtain

$$\sum_{m=1}^{M} \left(\mathrm{tr}(\mathbf{y}_m \mathbf{y}_m^H \mathbf{\Psi}_m \mathbf{R}_i \mathbf{\Psi}_m^H) - \sum_{l=1}^{L} \gamma_l \mathrm{tr}(\mathbf{\Psi}_m \mathbf{R}_l \mathbf{\Psi}_m^H \mathbf{\Psi}_m \mathbf{R}_i \mathbf{\Psi}_m^H) - \sigma^2 \mathrm{tr}(\mathbf{\Psi}_m \mathbf{\Sigma} \mathbf{\Psi}_m^H \mathbf{\Psi}_m \mathbf{R}_i \mathbf{\Psi}_m^H) \right) = 0.$$

Similarly, after taking the derivative with respect to σ^2, we obtain

$$\sum_{m=1}^{M} \left(\mathrm{tr}(\mathbf{y}_m \mathbf{y}_m^H \mathbf{\Psi}_m \mathbf{\Sigma} \mathbf{\Psi}_m^H) - \sum_{l=1}^{L} \gamma_l \mathrm{tr}(\mathbf{\Psi}_m \mathbf{R}_l \mathbf{\Psi}_m^H \mathbf{\Psi}_m \mathbf{\Sigma} \mathbf{\Psi}_m^H) - \sigma^2 \mathrm{tr}\left((\mathbf{\Psi}_m \mathbf{\Sigma} \mathbf{\Psi}_m^H)^2 \right) \right) = 0.$$

Therefore, we obtain $L+1$ equations for solving the $L+1$ optimization variables $(\gamma_1, \ldots, \gamma_L, \sigma^2)$, which can be written as the following system of equations

$$\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}_1^2}) & \ldots & \mathrm{tr}(\overline{\mathbf{R}_L \mathbf{R}_1}) & \mathrm{tr}(\overline{\mathbf{\Sigma} \mathbf{R}_1}) \\ \vdots & \ddots & \vdots & \vdots \\ \mathrm{tr}(\overline{\mathbf{R}_1 \mathbf{R}_L}) & \ldots & \mathrm{tr}(\overline{\mathbf{R}_L^2}) & \mathrm{tr}(\overline{\mathbf{\Sigma} \mathbf{R}_L}) \\ \mathrm{tr}(\overline{\mathbf{R}_1 \mathbf{\Sigma}}) & \ldots & \mathrm{tr}(\overline{\mathbf{R}_L \mathbf{\Sigma}}) & \mathrm{tr}(\overline{\mathbf{\Sigma}^2}) \end{pmatrix} \times \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_L \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(\mathbf{B}\mathbf{R}_1) \\ \vdots \\ \mathrm{tr}(\mathbf{B}\mathbf{R}_L) \\ \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}) \end{pmatrix}, \tag{59}$$

where

$$\mathrm{tr}(\overline{\mathbf{R}_i\mathbf{R}_j}) \doteq \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}(\boldsymbol{\Psi}_m\mathbf{R}_i\boldsymbol{\Psi}_m^H\boldsymbol{\Psi}_m\mathbf{R}_j\boldsymbol{\Psi}_m^H,)$$

computes the cross-correlation between distinct channels. Note that for $L = 1$, the system of equations reduces to the case of one primary user signal and noise levels, i.e., Equation (56). The former system of equations is governed by the system matrix (59). This matrix has as main diagonal the squared compressed normalized auto-correlation matrices at each frequency and for the noise term. The off-diagonal terms account for correlation among the distinct channel and the noise statistics. The higher the distinctness in terms of second-order statistics, the more "diagonal" the matrix (59) will be and the detector will operate with high accuracy. However, if the set $(\mathbf{R}_1, \ldots, \mathbf{R}_L, \boldsymbol{\Sigma})$ is highly correlated, the performance of the detector will be degraded.

Under the white noise assumption, the system of equations particularizes to, by fixing $\boldsymbol{\Sigma} = \mathbf{I}$

$$\begin{pmatrix} \mathrm{tr}(\overline{\mathbf{R}_1^2}) & \ldots & \mathrm{tr}(\overline{\mathbf{R}_L\mathbf{R}_1}) & K \\ \vdots & \ddots & \vdots & \vdots \\ \mathrm{tr}(\overline{\mathbf{R}_1\mathbf{R}_L}) & \ldots & \mathrm{tr}(\overline{\mathbf{R}_L^2}) & K \\ K & \ldots & K & K \end{pmatrix} \times \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_L \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(\mathbf{B}\mathbf{R}_1) \\ \vdots \\ \mathrm{tr}(\mathbf{B}\mathbf{R}_L) \\ \mathrm{tr}(\mathbf{B}) \end{pmatrix},$$

which for $L = 1$ reduces to a 2×2 system of equations and the solution is given by (58).

The multi-frequency model given by (37) encompasses many practical human-made communication signals, such as OFDM signals in DVB-T systems. In order to assess the performance in spectral analysis of the correlation-matching approaches proposed in this Section, we employ a toy example composed by an OFDM signal composed by 8 carriers with only 2 of which are activated.

We plot (see Figure 7) the spectral estimates of the OFDM signal with additive Gaussian noise (SNR of -10 dB), according to the Blackman-Tuckey and Capon estimates (11) and (15) with compressed-sampling and a compression rate of $1/2$ ($N = 2K$). As expected, the effect of noise enhancement appears in both estimators. Alternatively, we plot the power level estimates $(\gamma_1, \ldots, \gamma_8)$ according to the solution to the system of equations (59) (gray line). The compressed correlation-matching approaches is able to separate the white part of the noise from the compressed observations and, as a consequence, the power levels $(\gamma_1, \ldots, \gamma_8)$ accounts solely for the non-white part, i.e., the OFDM signal based on the normalized correlation matrices $\mathbf{R}_1, \ldots, \mathbf{R}_8$ following the model (37) with $\mathbf{C}_0$ a baseband OFDM signal generated according to the DVB-T in 2K-mode. The compressed correlation-matching estimate, hence, is a well suited approach for spectrum sensing detection in wideband (low-SNR) regimes.

4.3.3 Correlation-Matching for Stationary Signals

Assuming the knowledge on the normalized correlation lags is a robust assumption when having prior information on the communication systems employed by the primary users at a given bandwidth and location.

In this Section, we derive the optimal correlation-matching detector when there is no knowledge on the normalized signal correlation matrix. As $X(t)$ is a stationary signal, its correlation matrix is Toeplitz

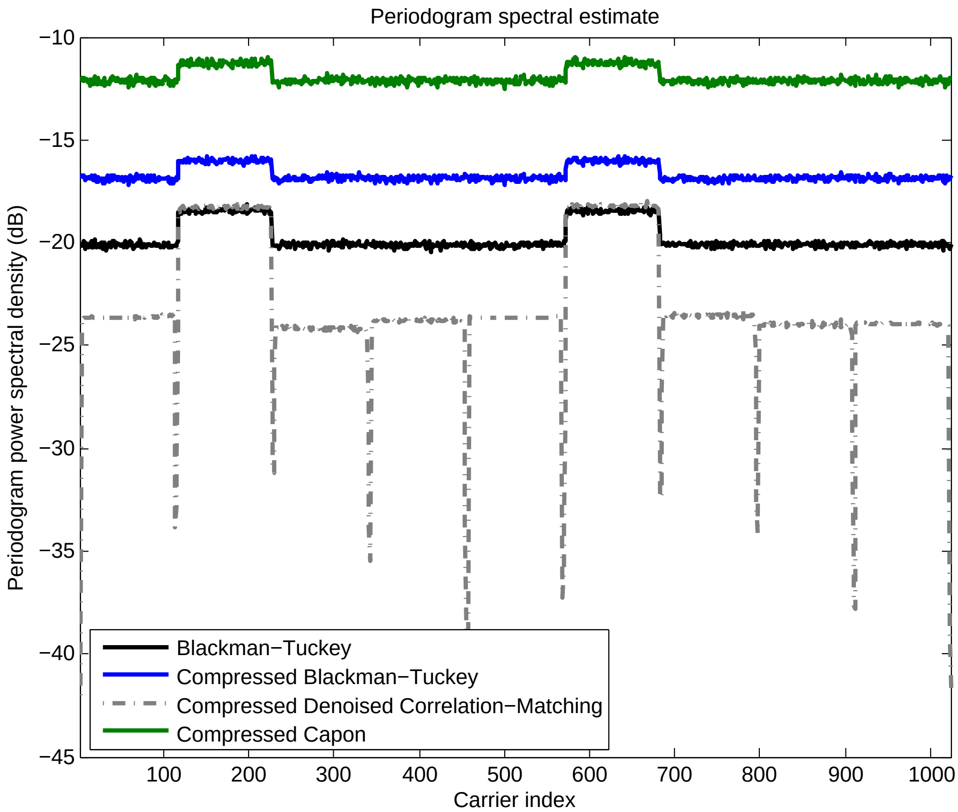

Figure 7: Spectral estimate of a $L = 8$ carriers OFDM signal with compression rate $1/2$.

and obeys the form

$$\mathbf{R}_X \doteq \begin{pmatrix} r[0] & r[-1] & \dots & r[-N+1] \\ r[1] & r[0] & \dots & r[-N+2] \\ \vdots & \vdots & \ddots & \vdots \\ r[-N+1] & \dots & r[1] & r[0] \end{pmatrix}. \tag{60}$$

The correlation-matching problem is then given by the solution to the N correlation lags and noise variance

$$\hat{\boldsymbol{r}}_X, \hat{\sigma}^2 = \arg\min_{\boldsymbol{r}_X, \sigma^2} \sum_{m=1}^{M} \left\| \mathbf{y}_m \mathbf{y}_m^H - \boldsymbol{\Psi}_m (\mathbf{R}_X + \sigma^2 \boldsymbol{\Sigma}) \boldsymbol{\Psi}_m^H \right\|^2,$$

where $\mathbf{R}_X$ has the structure as in (60). The compression matrix $\boldsymbol{\Psi}_m$ is constructed by randomly selecting K rows out of the $N \times N$ identity matrix. Let π_m denote the random set of rows selected by $\boldsymbol{\Psi}_m$. Then, $\boldsymbol{\Psi}_m$ will have a one in the $\pi(1)$-th column of the first row, a one in the $\pi_m(2)$-th column of the second row, up to a one in the $\pi_m(K)$-th column of the last row, with $1 \leq \pi_m(k) \leq N$. After some mathematical manipulations, it can be shown that the term $\boldsymbol{\Psi}_m \mathbf{R}_X \boldsymbol{\Psi}_m^H$ has the following structure

$$\left[\boldsymbol{\Psi}_m \mathbf{R}_X \boldsymbol{\Psi}_m^H\right]_{i,j} = \left[\mathbf{R}_X\right]_{\pi_m(i), \pi_m(j)} = r[\pi_m(i) - \pi_m(k)].$$

This formulation is not suitable for solving the problem (4.3.3) in a physical manner.

Alternatively, we propose an orthogonal decompostion for stationary correlation matrices based on a diagonal basis. Let $\tilde{\mathbf{D}}_n$ denote an all-zero matrix except an all-ones diagonal located n positions out of the main diagonal, e.g., for $N = 3$, $\tilde{\mathbf{D}}_0 = \mathrm{diag}(\ 1 \quad 1 \quad 1\ )$,

$$\tilde{\mathbf{D}}_{-2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \tilde{\mathbf{D}}_{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\tilde{\mathbf{D}}_{1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \text{ and } \tilde{\mathbf{D}}_{2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

This new formulation allows dealing with complex signals, as the signal correlation matrix can be expressed as $\mathbf{R}_X = \sum_{n=-N+1}^{N-1} \beta_n \mathbf{D}_n$, where $\mathbf{D}_n = \tilde{\mathbf{D}}_n \odot (\mathbf{s}(f)\mathbf{s}^H(f))$, with the Hermitian constraint $\beta_n = \beta_{-n}^*$. In such a case, the correlation-matching problem equals to solving the optimization problem

$$\begin{aligned} \hat{\boldsymbol{\beta}}, \hat{\sigma}^2 &= \arg\min_{\boldsymbol{\beta}, \sigma^2} \sum_{m=1}^{M} \left\| \mathbf{y}_m \mathbf{y}_m^H - \boldsymbol{\Psi}_m \left(\sum_{n=-N+1}^{N-1} \beta_n \mathbf{D}_n + \sigma^2 \boldsymbol{\Sigma} \right) \boldsymbol{\Psi}_m^H \right\|^2 \\ &= \arg\min_{\boldsymbol{\beta}, \sigma^2} \sum_{m=1}^{M} \left\| \mathbf{y}_m \mathbf{y}_m^H - \boldsymbol{\Psi}_m \left(\beta_0 \mathbf{D}_0 + \sum_{n=1}^{N-1} \left[\beta_n \mathbf{D}_n^H + \beta_n^* \mathbf{D}_n\right] + \sigma^2 \boldsymbol{\Sigma} \right) \boldsymbol{\Psi}_m^H \right\|^2, \end{aligned}$$

where $\boldsymbol{\beta} \doteq (\beta_0, \dots, \beta_{N-1})$ and we made use of the property $\mathbf{D}_n = \mathbf{D}_{-n}^H$. The former problem is convex on $\boldsymbol{\beta}$ and σ^2 and can be solved by taking the derivative with respect to the variabels and setting it to zero. On the one hand, we take the derivative with respect to β_i^*, $0 \leq i \leq N-1$, which after some mathematical manipulations, leads to the equation

$$\beta_0 \mathrm{tr}(\overline{\mathbf{D}_0 \mathbf{D}_i}) + \sum_{n=1}^{N} \left[\beta_n \mathrm{tr}(\overline{\mathbf{D}_n^H \mathbf{D}_i}) + \beta_n^* \mathrm{tr}(\overline{\mathbf{D}_n \mathbf{D}_i}) \right] + \sigma^2 \mathrm{tr}(\overline{\mathbf{\Sigma} \mathbf{D}_i}) = \mathrm{tr}(\mathbf{B}\mathbf{D}_i).$$

Similarly, after taking the derivative with respect to the noise level σ^2, we obtain the following equation,

$$\beta_0 \mathrm{tr}(\overline{\mathbf{D}_0 \mathbf{\Sigma}}) + \sum_{n=1}^{N} \left[\beta_n \mathrm{tr}(\overline{\mathbf{D}_n^H \mathbf{\Sigma}}) + \beta_n^* \mathrm{tr}(\overline{\mathbf{D}_n \mathbf{\Sigma}}) \right] + \sigma^2 \mathrm{tr}(\overline{\mathbf{\Sigma}^2}) = \mathrm{tr}(\mathbf{B}\mathbf{\Sigma}).$$

A suitable election on the correlation matrices basis $\{\mathbf{D}_n\}_{0 \leq n \leq N-1}$ is an orthogonal basis in the compressed domain, i.e., that $\mathrm{tr}(\mathbf{D}_i \mathbf{D}_j^H) \propto \delta_{ij}$. Indeed, the basis proposed in this section accomplishes the orthogonality property, and also $\mathrm{tr}(\mathbf{D}_i \mathbf{D}_j) \propto \delta_{i=j=0}$. In the compressed domain, the kernel of the scalar product is not preserved and, stochastically, some of the scalar products of the type $\mathrm{tr}(\overline{\mathbf{D}_i \mathbf{D}_i^H})$ might be zero, with low probability. Using this results, the derivative with respect to β_i^* is simplified to

$$\beta_i \mathrm{tr}(\overline{\mathbf{D}_i \mathbf{D}_i^H}) + \sigma^2 \mathrm{tr}(\overline{\mathbf{\Sigma} \mathbf{D}_i}) = \mathrm{tr}(\mathbf{B}\mathbf{D}_i), \tag{61}$$

from which we can obtain the ML estimate of $\boldsymbol{\beta}$ and the corresponding GLRT detector.

4.4 Mutual Information in Compressed-Sampling

We next discuss the design of the compression matrices $\mathbf{\Psi}_m$, $1 \leq m \leq M$. The objective of $\mathbf{\Psi}_m$ is to compress or project the sparse signal onto a space of smaller dimension such that the information contained in it is sufficient to recover $\mathbf{x}$ or $S_X(f)$. The compression matrix in literature is usually chosen to be a puncturing or pinning matrix, which randomly selects K out of N samples of the original signal. One advantage of this choice is that the compression matrix has the property $\mathbf{\Psi}_m^H \mathbf{\Psi}_m = \mathbf{I}$. Another possibility is to use Gaussian random entries to build the compression matrix. In such a case, the product $\mathbf{\Psi}_m^H \mathbf{\Psi}_m$ asymptotically goes to the identity matrix as the size of the observation grows.

4.4.1 Capacity Analysis

From an information-theoretical perspective, the set $\{\mathbf{\Psi}_m\}$ can be casted as M multiple-input multiple-output channels with N inputs and K outputs, where the same statistical input, given by the correlation matrix $\mathbf{R}_X$, is injected to the channels. Consider now that $\mathbf{\Psi}$ is fixed for the M channels. Let $\tilde{\mathbf{x}}_m = \mathbf{x}_m + \mathbf{w}_m$ denote the noisy observations of a sparse wideband spectrum. The compressed-sampling framework for fixed compression matrix is depicted in Figure 8a.

Typically, the achievable capacity is then given by assuming Gaussian-coded input, i.e.,

$$\mathcal{C}(\mathbf{\Psi}) \doteq \max I(Y;X) = \log\det\left(\frac{\mathbf{\Psi}(\mathbf{R}_X + \mathbf{R}_W)\mathbf{\Psi}^H}{\mathbf{\Psi}\mathbf{R}_W\mathbf{\Psi}^H}\right), \tag{62}$$

where we made use of the accepted notation $\det(\mathbf{A}/\mathbf{B}) = \det(\mathbf{A}\mathbf{B}^{-1}) = \det(\mathbf{B}^{-1}\mathbf{A})$, and where $\mathbf{R}_W \doteq \mathbb{E}[\mathbf{w}\mathbf{w}^H]$ stands for the noise plus interference covariance matrix. By the general properties of the com-

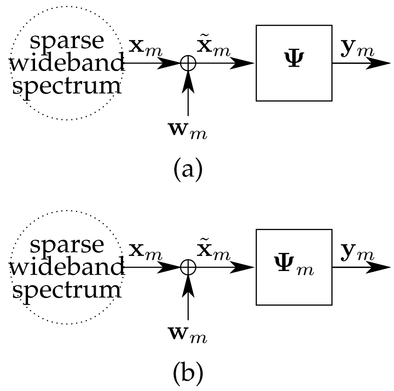

Figure 8: Compressed-sampling framework with fixed (a) and variable (b) compression matrix.

pression matrix, we note that the inverse $(\boldsymbol{\Psi}\mathbf{R}_W\boldsymbol{\Psi}^H)^{-1}$ can be expressed as $\boldsymbol{\Psi}^{-H}\mathbf{R}_W^{-1}\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}\mathbf{R}_W^{-1}\boldsymbol{\Psi}^H$. This observation allows the capacity (62) to further be reduced to

$$\begin{aligned} \mathcal{C}(\boldsymbol{\Psi}) &= \log\det\left(\mathbf{I} + \boldsymbol{\Psi}\mathbf{R}_W^{-1}\boldsymbol{\Psi}^H\boldsymbol{\Psi}\mathbf{R}_X\boldsymbol{\Psi}^H\right) && \text{(63a)} \\ &\approx \begin{cases} \operatorname{tr}\left(\boldsymbol{\Psi}\mathbf{R}_W^{-1}\boldsymbol{\Psi}^H\boldsymbol{\Psi}\mathbf{R}_X\boldsymbol{\Psi}^H\right) & \text{wideband} \\ \log\det\left(\boldsymbol{\Psi}\mathbf{R}_W^{-1}\boldsymbol{\Psi}^H\boldsymbol{\Psi}\mathbf{R}_X\boldsymbol{\Psi}^H\right) & \text{narrowband,} \end{cases} && \text{(63b)} \end{aligned}$$

where the approximation has been done under the wideband or low-SNR, and narrowband or hight-SNR assumptions.

Conversely, consider that the compression is in general dependent on the realization m, that is, M random process independent realizations as depicted in Figure 8b. Even though the achievable capacity would be given by the conditional mean, here we assume that the set $\{\boldsymbol{\Psi}_m\}$ is known by the cognitive radio user. In such a case, we define the achievable channel capacity as [93]

$$\mathcal{C}(\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_M) \doteq \frac{1}{M}\sum_{m=1}^{M}\log\det\left(\frac{\boldsymbol{\Psi}_m(\mathbf{R}_X+\mathbf{R}_W)\boldsymbol{\Psi}_m^H}{\boldsymbol{\Psi}_m\mathbf{R}_W\boldsymbol{\Psi}_m^H}\right), \quad (64)$$

which can be also simplified using the properties of the compression matrix to the following expressions under the wideband and narrowband regimes:

$$\begin{aligned} \mathcal{C}(\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_M) &= \frac{1}{M}\sum_{m=1}^{M}\log\det\left(\mathbf{I} + \boldsymbol{\Psi}_m\mathbf{R}_W^{-1}\boldsymbol{\Psi}_m^H\boldsymbol{\Psi}_m\mathbf{R}_X\boldsymbol{\Psi}_m^H\right) && \text{(65a)} \\ &\approx \begin{cases} \frac{1}{M}\sum_{m=1}^{M}\operatorname{tr}\left(\boldsymbol{\Psi}_m\mathbf{R}_W^{-1}\boldsymbol{\Psi}_m^H\boldsymbol{\Psi}_m\mathbf{R}_X\boldsymbol{\Psi}_m^H\right) & \text{wideband} \\ \frac{1}{M}\sum_{m=1}^{M}\log\det\left(\boldsymbol{\Psi}_m\mathbf{R}_W^{-1}\boldsymbol{\Psi}_m^H\boldsymbol{\Psi}_m\mathbf{R}_X\boldsymbol{\Psi}_m^H\right) & \text{narrowband} \end{cases} . && \text{(65b)} \end{aligned}$$

The formulation presented in [93] is suitable for the problem in hand because, in the limit $M \to \infty$, (64) becomes the expectation, that is,

$$\lim_{M\to\infty}\mathcal{C}(\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_M) = \mathbb{E}_{\Psi}\left[\mathcal{C}(\Psi)\right] = \mathbb{E}_{\Psi}\left[\log\det\left(\frac{\Psi(\mathbf{R}_X+\mathbf{R}_W)\Psi^H}{\Psi\mathbf{R}_W\Psi^H}\right)\right],$$

where Ψ denotes the matrix random variable from which the set of compression matrices $(\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_M)$ is obtained. When the compression matrices are chosen as selection matrices, (65) is simplified to the following under the wideband and narrowband regimes:

$$\begin{aligned} \mathbb{E}_\Psi\left[\mathcal{C}(\Psi)\right] &= \mathbb{E}_\Psi\left[\log\det\left(\mathbf{I} + \Psi\mathbf{R}_W^{-1}\Psi^H\Psi\mathbf{R}_X\Psi^H\right)\right] & \text{(66a)} \\ &\approx \begin{cases} \mathbb{E}_\Psi\left[\operatorname{tr}\left(\Psi\mathbf{R}_W^{-1}\Psi^H\Psi\mathbf{R}_X\Psi^H\right)\right] & \text{wideband} \\ \mathbb{E}_\Psi\left[\log\det\left(\Psi\mathbf{R}_W^{-1}\Psi^H\Psi\mathbf{R}_X\Psi^H\right)\right] & \text{narrowband} \end{cases}. & \text{(66b)} \end{aligned}$$

From an analytical perspective, when the compression matrix is fixed for all the observations, the same reference samples are taken to compute the sufficient statistics for the sensing problem. That is, the compressed sample covariance matrix common in all expressions in (63), i.e., $\mathbf{\Psi}\mathbf{R}_X\mathbf{\Psi}^H$ selects K rows and K columns out of $\mathbf{R}_X$. On the other hand, when the compression matrix is variable for each sample or for each subset of samples, then the overall capacity is a random process whose expected value is given in (66). It is easy to check that the product $\Psi^H\Psi$ randomly places K ones at the main diagonal uniformly. Hence, we have $\mathbb{P}\left([\Psi^H\Psi]_{ij} = 1\right) = K/N\delta_{ij}$, where δ_{ij} is the Kronecker delta. Then, we show that

$$\mathbb{E}_\Psi[\Psi^H\Psi] = \frac{K}{N}\mathbf{I}.$$

Therefore, it is expected that the capacities in (66) achieve the full capacity of a non-compression system scaled by a factor K/N. However, the capacities in (63) depend on which eigenmodes of the signal and noise covariance matrices are selected. For large data records, i.e., when the size of the observation asymptotically grows with $N \to \infty$, the eigenvalues and eigenvectors of $\mathbf{R}_X$ and $\mathbf{R}_W$ approximate to the spectra [94]

$$\begin{gathered} \lambda_n(\mathbf{R}_X) = S_X(f_n), \quad \lambda_n(\mathbf{R}_W) = S_W(f_n) \\ \mathbf{u}_n = \tfrac{1}{\sqrt{N}}\left[1 \ \exp(\jmath 2\pi f_n) \ \ldots \ \exp(\jmath 2\pi (N-1) f_n)\right]^T \end{gathered}$$

where $S_X(f_n)$ and $S_W(f_n)$ denotes the power spectral density of the observations and the nosie at $f_n \doteq n/N$, respectively. Under this assumption, (63) can be written as

$$\mathcal{C}(\mathbf{\Psi}) \approx \begin{cases} \operatorname{tr}\left(\mathbf{\Psi}\mathbf{R}_W^{-1}\mathbf{\Psi}^H\mathbf{\Psi}\mathbf{R}_X\mathbf{\Psi}^H\right) = \sum_{n\in\mathcal{K}} \frac{S_X(f_n)}{S_W(f_n)} & \text{wideband} \\ \log\det\left(\mathbf{\Psi}\mathbf{R}_W^{-1}\mathbf{\Psi}^H\mathbf{\Psi}\mathbf{R}_X\mathbf{\Psi}^H\right) = \sum_{n\in\mathcal{K}} \log\frac{S_X(f_n)}{S_W(f_n)} & \text{narrowband,} \end{cases},$$

where as, similarly, the capacities (66) for large data records become

$$\mathbb{E}_\Psi\left[\mathcal{C}(\Psi)\right] \approx \begin{cases} \mathbb{E}_\Psi\left[\operatorname{tr}\left(\Psi\mathbf{R}_W^{-1}\Psi^H\Psi\mathbf{R}_X\Psi^H\right)\right] = \frac{K}{N}\sum_{n=1}^{N}\frac{S_X(f_n)}{S_W(f_n)} & \text{wideband} \\ \mathbb{E}_\Psi\left[\log\det\left(\Psi\mathbf{R}_W^{-1}\Psi^H\Psi\mathbf{R}_X\Psi^H\right)\right] = \frac{K}{N}\sum_{n\in\mathcal{K}}\log\frac{S_X(f_n)}{S_W(f_n)} & \text{narrowband} \end{cases}.$$

Here, we have defined $\mathcal{K}$ as the location of the K ones in the main diagonal of the product $\mathbf{\Psi}^H\mathbf{\Psi}$.

This has become a typical problem of deviation: while for fixed compression matrix we obtain a capacity given by the $\mathbf{\Psi}$ realization, for variable compression matrix we expect the average capacity, which is the capacity in the case of no compression scaled by the compression rate. Therefore, we need to analyze the former capacity expressions from an outage probability viewpoint. In other words, evaluate $\mathbb{P}\left(\mathcal{C}(\Psi) \geq \mathbb{E}_{\Psi}\left[\mathcal{C}(\Psi)\right]\right)$.

In an sparse wideband environment, the probability that $\mathcal{S}$ selects the eigenmodes that contain primary user signal, with large compression rate, is low. Moreover, it is a fundamental limit that in the best case, it selects all the non-zero eigenmodes, hence the wideband capacity equals the no-compression capacity. Let σ_X denote the sparsity set of the wideband signal X when the size of the observation is N, i.e, $S_X(f_n) > 0$ for $|\sigma_X| = S$ values of n, and zero otherwise. Then,

$$\mathbb{E}[\mathcal{C}(\mathbf{\Psi})] \approx \frac{K}{N} \sum_{n \in \sigma_X} \mathrm{SNR}_n,$$

An alternative manner to evaluate the effect of the compression matrix in the process of recovery is making use of the concept of outage probability. Outage probability evaluates the probability that some performance indicator falls below or above a given value corresponding to the maximum or minimum allowable, respectively. Even though the purpose in this work is distinct, the outage probability has been widely used in evaluating the performance of digital communication systems from a wide variety of perspectives [60, 61, 63–67, 95].

The outage probability of the process of data reconstruction is given by

$$\mathrm{OP}(\mathbf{\Psi}) = \mathbb{P}(\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \leq \epsilon), \tag{67}$$

where ϵ defines the reconstruction accuracy in the mean-squared error sense.

4.5 Medium Access in Cognitive Radio

In order to provide an assessment on the convenient MATs for cognitive radio, we consider two independent approaches from an information-theoretical perspective. On the one hand, the classical MAC capacity will be analyzed when employing the proposed MATs. On the other hand, in a more pragmatic step, the analysis will focus on the so-called transmission capacity, where the target is to maximize the geographical density of reliable transmissions.

4.5.1 Classical Information-Theoretical Perspective

In this first approach, the well known achievable Ahlswede-Liao capacity region [96, 97] will be considered. Before a more concrete presentation, the following itemized remarks are relevant.

- The *achievable MAC capacity* is based on infinite delay latency, i.e., it is only an asymptotic formulation of the problem, making the analysis of partial practical interest.
- The analysis becomes useful in the comparison of basic MAC schemes, as for TDMA, MF-TDMA, and CDMA, being possible to state that some of MATs cannot provide working conditions achieving the entire possible capacity region.
- In the MAC problem, the optimal resources assignments are not unique, that is, the optimization of the rates, assignments to the different users can be performed under many different fairness approaches. Typically, the most common characterizations are:

- The *single-user rates*, R_k^1 are the maximum achievable rates based on multiple single-user communication scenarios. This is still very common approach, and it is coexisting with multi-used detection and multiple-access interference mitigation enhanced methods. The single-user rates are an indicator of te achievable rates of each user and, therefore, of each user's channel capacity. In other words, if the system ensures a certain rate R_k to the k-th user in the MAC, the ratio R_k/R_k^1 represents the relative cost that the systems pays due to other users sharing the resources, a parameter which is more representative than the absolute value.
- The *maximum sum-rate* corresponds to maximize the total rate of the wireless network, i.e., $\sum_k \alpha_k R_k$ subject to $\sum_k \alpha_k = 1$, where the parameters α_k establish the priorities for the users. In a non-weighted scenario, $\alpha_k = 1/K$ and the maximization is done over the direct sum of rates, $\sum_k R_k$. This has been widely used in literature, but might result in unbalanced rate-allocations to different users, depending on the channel propagation conditions [98–100].
- The *maximum common rate*, which is obtained by setting $R_1 = \ldots = R_K$. This approach under performs when the single-user rates are very diverse, as the common rate is a waist of resources because if forces the users with the best channels to lower their rate dramatically to reach the level of the weakest channels.
- When establishing a *set of rate requirements*, denoted by $\{\varphi_1, \ldots, \varphi_K\}$, the system is optimized to achieve $R_k = \varphi_k$ for each user. This structure assumes there is enough available power at the transmitter side and that the optimization process equals to find the minimum power to achieve the rate requirement (e.g. [101]). With sum power constraint, this scheme might not be possible —maybe only a small number of users are served— and some other strategies are considered, like
- the *maximum balanced-rate*, first explored in [102], is such that

  $$\frac{R_1}{\varphi_1} = \ldots = \frac{R_K}{\varphi_K},$$

  where φ_k represents the target rate for user k, and it is usually set to $\varphi_k = R_k^1$. This means that all the users transmit at a rate proportional to the single-user rate offered by their own channel. As a result, the relative cost implied by the coexistence with other users in the same system is the same for all users. Though it seems to be the most fair situation, it has only been analyzed in [102] for the scalar memoryless broadcast providing water-filling algorithm, and [103] for multi-tone systems, providing loading for the two users case. This approach is useful when individual rate requirements are crucial.
- The *maximum rate-product* considers the maximization of $\Pi_k R_k^{\tilde{\alpha}_k}$ where now the priorities are normalized to $\Pi_k \tilde{\alpha}_k = 1$. This case ensures that if a user is unfairly treated, it is reflected with more weight in the target function and, as a result, the system will tend to treat all the users more equally, taking into account their priorities. It is worth noting that the maximum rate-product becomes equivalent to the maximum common-rate if there are no additional design constraints in the overall system optimization and $\alpha_k = 1$. It is more evident that in this criterion the rate-allocation will tend to be equalized.

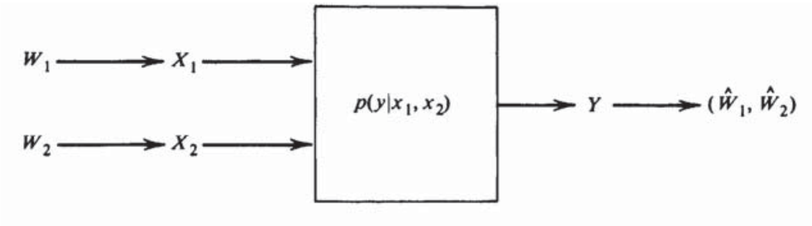

Figure 9: The Multiple-Access Channel with $M = 2$ users.

- Conventionally, the MAC management requires of a centralized control scheme. It is nowadays possibe to start considering distributed cooperative-network management approaches, with the different access-nodes cooperating in the dynamic resource allocations.
- Considering achievable transmission rates is of interest in the definition of meduim to long-term standards as the most advanced known coded-modulation schemes become feasible with the current and future computation capabilities.

For the sake of clarity, the presentation of the MAC capacity will be performed in the simplest case with two access terminals, according to the model in which each user encodes the information messages W_1 and W_2, and transmits the coded information X_1 and X_2 over a multiple-access communication channel characterized by the transitions probabilities $P(Y|X_1, X_2)$. The output of the channel, Y, is then decoded to recover the original information of the two users, $\hat{W}_1$ and $\hat{W}_2$. Let $\mathbf{x}_m = (x_1^m, \ldots, x_N^m)$ be N consecutive encoded symbols of user m. For a memoryless MAC, the following input-output characterization will apply

$$\mathbb{P}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \prod_{n=1}^{N} p(y_n|x_n^1, x_n^2).$$

Accordingly, the Ahlswede-Liao capacity region (R_1, R_2) is given by the following set of equations [96]:

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) \\ R_2 &\leq I(X_2; Y|X_1) \\ R_1 + R_2 &\leq I(X_1, X_2; Y). \end{aligned} \tag{68}$$

The normal (non-degenerated) case for the three constraints is depicted in Figure 10 [14], where the pair (R_1, R_2) is achievable inside the convex hull set.

TDMA and MF-TDMA MATs. In the TDMA MAC access, different users are allocated to different disjoint time slots. In the same manner, for MF-TDMA MAC access, the allocation is performed at disjoint time-frequency slots. These are two classical MATs, and are based on coded-modulations not well suited for a multi-user detection (MUD) at the receiver in the uplink. The difficulty in the adoption of feasible MUD is mostly because of the complexity in the joint synchronization of the signals and not because of the joint detection and decoding, which is indeed possible. For these two cases, we have that

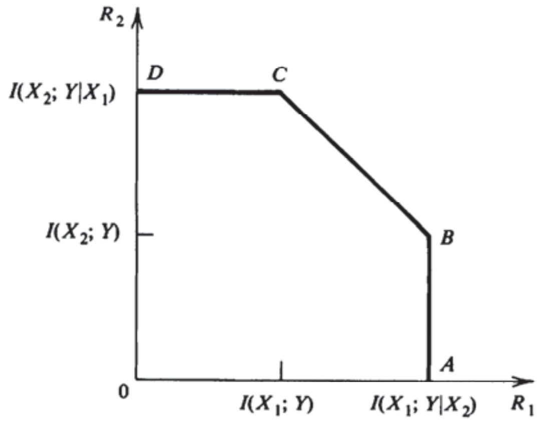

Figure 10: Capacity Region of the Multiple-Access Channel with $M = 2$ users.

the Ahlswede-Liao capacity region (68) particularizes to

$$\begin{aligned} R_1 &\leq I(X_1; Y) \\ R_2 &\leq I(X_2; Y) \\ R_1 + R_2 &\leq \alpha I(X_1; Y) + (1-\alpha) I(X_2|Y), \end{aligned} \tag{69}$$

where the parameter $0 \leq \alpha \leq 1$ stands for the percentage of channel occupation of the first user, while $(1-\alpha)$ is for the second user. The capacity region is then according to a convex hull given by (...).

The Gaussian MAC. In this section, we will make use of the Gaussian encoding theory, such that the ultimate achievable performance for any system is characterized. The signal model is then a memoryless Gaussian MAC scheme, i.e., the received signal at the uplink is given by $Y = X_1 + X_2 + Z$, where Z denotes the additive Gaussian noise of variance $\mathbb{E}(|Z|^2) = \sigma_Z^2$. Here, hence, we assume that both user 1 and 2 make use of Gaussian codes. Any MAC will be upper-bounded by the following three equations:

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2) \leq \tfrac{1}{2} \log \left(1 + \tfrac{\sigma_{X_1}^2}{\sigma_Z^2}\right) \\ R_2 &\leq I(X_2; Y|X_1) \leq \tfrac{1}{2} \log \left(1 + \tfrac{\sigma_{X_2}^2}{\sigma_Z^2}\right) \\ R_1 + R_2 &\leq I(X_1, X_2; Y) \leq \tfrac{1}{2} \log \left(1 + \tfrac{\sigma_{X_1}^2 + \sigma_{X_2}^2}{\sigma_Z^2}\right), \end{aligned} \tag{70}$$

where $\sigma_{X_m}^2$ denotes the received power of signals X_m, $1 \leq m \leq M$. It is worth noting that the upper bounds are achievable for Gaussian codes when the information of users 1 and 2 are statistical independent, i.e., when the three inequalities in (70) become equalities. It is also remarkable that the expressions are the one-dimension capacities, and the scaling factor could be removed for an alternative two-dimensional formulation.

Despite the general analysis is possible, two asymptotic cases will be considered for clarity, i.e., the analysis in the high-SNR and wideband regimes.

The Wideband Regime. Let's define the two SNRs $\gamma_1 \doteq \sigma^2_{X_1}/\sigma^2_Z$ and $\gamma_2 \doteq \sigma^2_{X_2}/\sigma^2_Z$. In this case of study, we will assume that both users work in the wideband regime. It is well known that the wideband regime is characterized by low SNRs and close to zero spectral efficiency [11]. Therefore, we can make the low-SNR approximations $\log(1+\gamma_1) \approx \gamma_1$, $\log(1+\gamma_2) \approx \gamma_2$ and, similarly, $\log(1+\gamma_1+\gamma_2) \approx \gamma_1+\gamma_2$. Hence, the capacity region becomes

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2) \approx \tfrac{1}{2}\left(\frac{\sigma^2_{X_1}}{\sigma^2_Z}\right) \\ R_2 &\leq I(X_2;Y|X_1) \leq \tfrac{1}{2}\left(\frac{\sigma^2_{X_2}}{\sigma^2_Z}\right) \\ R_1+R_2 &\leq I(X_1,X_2;Y) \leq \tfrac{1}{2}\left(\frac{\sigma^2_{X_1}+\sigma^2_{X_2}}{\sigma^2_Z}\right), \end{aligned} \tag{71}$$

which, in fact, correspond to a rectangular region delimited by $R_1 \leq \frac{1}{2}\gamma_1$ and $R_2 \leq \frac{1}{2}\gamma_2$. This result can be easily generalized to a higher number of users, e.g., M. The conclusion is that the achievable rates for the different users are only limited by their own SNRs., and there is not impact on the multiple-access interference. This analysis suggests that the noise contribution is dominant and a spread-spectrum strategy with an optimal multiple-access interference cancellation for DS-CDMA or FH-CDMA alternative with a null hopping collision.

The High-SNR Regime. On the other hand, this second assumption is intended for characterizing the MF-TDMA MAT. We will assume that the SNRs are arbitrarely large such that the following approximations are fulfilled on the achievable rate capacity:

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2) \approx \tfrac{1}{2}\log\left(\frac{\sigma^2_{X_1}}{\sigma^2_Z}\right) \\ R_2 &\leq I(X_2;Y|X_1) \leq \tfrac{1}{2}\log\left(\frac{\sigma^2_{X_2}}{\sigma^2_Z}\right) \\ R_1+R_2 &\leq I(X_1,X_2;Y) \leq \tfrac{1}{2}\log\left(\frac{\sigma^2_{X_1}+\sigma^2_{X_2}}{\sigma^2_Z}\right). \end{aligned} \tag{72}$$

Despite is not possible to go further on equations (72), we sill now focuss on the case there is no near-far between users, that is, for $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2_X$. We then have that the inequalities in (72) all become $R_1 \leq \frac{1}{2}\log(1+\sigma^2_X/\sigma^2_Z)$, $R_2 \leq \frac{1}{2}\log(1+\sigma^2_X/\sigma^2_Z)$, and $R_1+R_2 \leq \frac{1}{2}\log(1+\sigma^2_X/\sigma^2_Z)$, being the latter the most restrictive. We conclude then that TDMA becomes optimal in the case of a high-SNR regime without near-far effect between users.

4.5.2 A Novel Outage Perspective: Transmission Capacity

We now discuss the baseline formulation for the transmission capacity in the MAC of wireless networks, and provide upper and lower bounds for sake of mathematical tractability. We then discuss the effect of the Rayleigh channel and derive the corresponding bounds the MATs under assessment.

Fundamentals on Transmission Capacity. We first consider the simpler case where the propagation conditions are given exclusively in terms of distance between terminals and the path-loss exponent n, for $n \geq 2$. We assume that a set of M terminals are geographically distributed according to a Poisson point process (PPP) of intensity λ —this roughly means that there are λ terminals per unit area—, whose locations are given by the set $\Pi(\lambda) \subset (\mathbb{R}^2)^M$.

The transmission capacity is defined as the spatial intensity of attempted transmissions associated to an outage probability of the SINR. The outage probability (OP), denoted by q, is the probability that the SINR at the MAC receiver is below a specified level β required for reliable communication, i.e.,

$$q(\lambda) \doteq \mathbb{P}\left(\mathrm{SINR} < \beta\right), \tag{73}$$

where the expression of the SINR depends on the MAT and transmit and receive signal processing used in the wireless network. The randomness of the SINR is on the interferer locations $\Pi(\lambda)$, and the OP is a function of n, β, and λ. Note that $0 \leq (\lambda) \leq 1$ is a continuous monotone function increasing in λ. The transmission capacity is written as

$$c(\epsilon) \doteq \arg_\lambda q(\lambda) \leq \epsilon. \tag{74}$$

In (74), the quantity ϵ denotes the outage probability of $\mathbb{P}(\mathrm{SINR} < \beta)$, and is a network-wide quality of service measure, ensuring a typical attempted transmission will succeed with probability $(1 - \epsilon)$. Note that the transmission capacity is a valid metric for MATs that have multiple-access interference (such as CDMA or SDMA), whereas multiple-access techniques in which users have disjoint allocated resources (such as TDMA or MF-TDMA) the equivalent SINR becomes the SNR, hence independent on λ.

Bounds on Transmission Capacity. The exact formulation for $c(\epsilon)$ can only be obtained for some particular cases, such as path-loss only propagation with $n = 4$, or Rayleigh fading channel. The main reason for the unavailability of a closed-form solution of (73) and (74) lies in the essential difficulty in computing the distribution of the SINR, which as mentioned above, depends on the interferer locations and the parameters λ, n, and β. This motivates the search of lower and upper bounds for the transmission capacity [59], which are discussed next.

A lower bound on the probability of outage is obtained by partitioning the set of interferers $\Pi(\lambda)$ into dominating and non-dominating nodes, i.e., near and far nodes. Here, a node is considered to be dominant if its interference contribution alone is sufficient to cause outage at the MAC receiver. Consider any dominant node X_m, located at distance r_m to the MAC receiver. The individual SINR of a reference transmitter located at distance r is given by

$$\mathrm{SINR} \doteq \frac{\sigma_X^2}{\sigma_Z^2 + \sigma_{X_m}^2},$$

where $\sigma_X^2 = \rho r^{-n}$, $\sigma_{X_m}^2 = \rho r_m^{-n}$, and σ_Z^2 is the noise variance. The parameter ρ reflects the transmitted power, which for sake of clarity, we assume the same for all the terminals. The set of dominant nodes, which we denote by $\Pi_1(\lambda)$, is then equivalent to

$$\Pi_1(\lambda) = \{r_m \ : \ \mathrm{SINR} < \beta\} = \left\{r_m \ : \ \frac{\rho r^{-n}}{\sigma_Z^2 + \rho r_m^{-n}} < \beta\right\} \equiv \left\{r_m \ : \ r_m < \kappa^{-1/n}\right\},$$

where $\kappa \doteq r^{-n}/\beta - \sigma_Z^2/\rho$. In other words, the set of dominant interferers is geographically defined as the interferers located inside a disk of radius $\kappa^{-1/n}$. By ignoring the far nodes, the lower bound on (73) is given by the probability that an interferer is inside the disk of radius $\kappa^{-1/n}$. Since terminals are distributed according to a PPP with spatial intensity λ, the outage probability (73) becomes $q(\lambda) \geq 1 - \exp\left(-\lambda\pi(\kappa^{-1/n})^2\right)$, that is $1 - \mathbb{P}(r \leq \kappa^{-1/n})$. This upper-bound on the outage probability allows to determine the intensity λ to satisfy the relation (74), which is given by

$$\begin{aligned} c(\epsilon) \leq \lambda^* &= \arg_\lambda \left\{1 - \exp\left(-\lambda\pi(\kappa^{-1/n})^2\right) = \epsilon\right\} \\ &= -\frac{\log(1-\epsilon)}{\pi}\kappa^{2/n}. \end{aligned}$$

As the outage ϵ is small, we can further simplify the expression for the tranmssion capacity by making use of the inequality $-\log(1-\epsilon) \leq \epsilon$. This allows the transmission capacity to be upper bounded by [59]

$$c(\epsilon) \leq \frac{\epsilon}{\pi}\kappa^{2/n}. \tag{75}$$

Similarly, an upper bound on (73) and the corresponding lower bound on (74) can be obtained by only considering the far terminals. Using the Chebychev inequality [59], it can be stated that

$$c(\epsilon) \geq \frac{1}{2}\frac{\epsilon}{\pi}\kappa^{2/n}. \tag{76}$$

The tightness of the bounds is analyzed in [59], which concludes that the upper bound (75) is much tighter than the upper bound (76). Therefore, in the following assessment of the FH-CDMA and DS-CDMA multiple-access methods, we focus on the upper bound (75).

4.5.3 Spread-Spectrum Multiple-Access

In this section, we discuss the transport capacity for the FH-CDMA and DS-CDMA methods with spreading factor parameter K.

FH-CDMA. Frequency-hopping CDMA is a MAT that exploits the frequency diversity and the randomness of the carrier selection to enhance the robustness in front of channel fading and multiple-access interference, and has important inherent security features. A FH-CDMA system divides the total bandwidth B into a large number of non-overlapping frequency slots or carriers of bandwidth B/K, being K the number of bands. In any signaling interval, the transmitted signal occupies one or more of the available frequency slots. The selection of the frequency slots in each signal interval is made pseudo-randomly, according to a pseudo-random sequence. Assuming perfect synchronization with the receiver, the pseudo-random frequency translation introduced at the transmission is successfully removed at the demodulator, and the data signal is extracted —usually a FSK modulation with non-coherent demodulation is employed.

The frequency-hopping rate can be selected equal, lower, or higher than the symbol rate. If the rate is equal or lower than the symbol rate, the FH-CDMA is then called a slow-hopping system; whereas is the hopping rate is higher, i.e., there are multiple frequency hops per symbol, the system is a fast-hopping CDMA. However, there is a penalty in subdividing an information symbol into several frequency-hopped elements, because the energy from the separate elements is combined non-coherently.

By its nature, FH-CDMA provides a high probability of interference avoidance, which grows with the number of frequency slots K. Specifically, if there are other users sharing the frequency bandwidth, there are occasional collisions when two transmitters select the same frequency slots, causing mutual multiple-access interference. These collisions happen with low probability, and it is then possible to recover from a moderate number of collisions when time-coding is employed. For this particular reason, the FH-CDMA MAT is known to achieve a high capacity even in dense wireless networks.

We next state the transmission capacity upper-bound of FH-CDMA in a wireless network where the terminals are PPP geographically distributed.

Proposition 5. *The transport capacity for a FH-CDMA system with K carriers and outage probability ϵ is given by the upper-bound [67]*

$$c_{FH}(\epsilon) \leq \frac{\epsilon K}{\pi} \kappa^{2/n}, \tag{77}$$

where $\kappa \doteq r^{-n}/\beta - \sigma_Z^2/\rho$.

Proof. Let β be the nominal SINR requirement for FH-CDMA. An arbitrary receiver on a wireless network using FH-CDMA only sees interference from transmitters on subcarrier k, and the aggregate interference at each subcarrier must be such that the SINR exceeds the threshold β. Because of the independent sampling assumption, each process $\Pi_k(\lambda_k)$ is an homogeneous Poisson point process with intensity $\lambda_k = \lambda/K$, $1 \leq k \leq K$. The outage probability (73) is then given by

$$q_{FH}(\lambda_k) = \mathbb{P}\left(\frac{\rho r^{-n}}{\sigma_Z^2 + \sum_{l \in \Pi_k(\lambda_k)} \rho r_l^{-n}} < \beta\right),$$

for $1 \leq k \leq K$. The noise power $\sigma_Z^2 = N_0 \frac{B}{K}$ accounts for the active k-th subcarrier, being N_0 the noise power spectral density, and B the total bandwidth. From the upper bound (75), we have that $\lambda_k = \frac{\lambda}{K} \leq \frac{\epsilon}{\pi}\kappa^{2/n}$, or, equivalently, $c_{FH}(\epsilon) \leq \frac{\epsilon K}{\pi}\kappa^{2/n}$. □

DS-CDMA. Direct-sequence CDMA (DS-CDMA) allows several users to simultaneously access a given frequency allocation by spreading the modulated waveform of narrow bandwidth B/K over the wider bandwidth B using spreading codes unique for each user, by a factor of K. DS-CDMA spreads the signal by directly multiplying the data waveform with a high bandwidth pseudo-random sequence. Assuming perfect synchronization, the receiver is then able to detect the transmitted signal because the noise and the interference signals become attenuated by a factor of approximately K.

However, because of other users do not use completely orthogonal spreading codes, there is a residual multiple-access interference present at the multi-user demodulator, which can be a critical problem if the SINR is low. Residual interference is usually overcome at higher layers with multiple-access control, which along with power allocation can effectively perform spectrum efficiency maximization and interference mitigation [104]. One way to further improve the performance of FH-CDMA is to use spectrum sensing to avoid collisions, which consist of sensing the hopping sequences of the nearby terminals. The spectral efficiety of DS-CDMA has been analyzed by Verdu and Shamai in [105], whereas the transmission capacity is given by the following statement.

Proposition 6. *The transport capacity for a DS-CDMA system with a spreading factor K and outage probability ϵ is given by the upper-bound [67]*

$$c_{DS}(\epsilon) \leq \frac{\epsilon}{\pi}(\kappa K)^{2/n}, \tag{78}$$

where $\kappa \doteq r^{-n}/\beta - \sigma_Z^2/\rho$.

Proof. Contrarily to FH-CDMA, DS-CDMA uses the spreading factor K to reduce the minimum SINR required for reliable communication. If the nominal threshold is β, then DS-CDMA reduces the SINR requirement to β/K, assuming a typical pseudo-noise code [106]. The total noise power in DS-CDMA is $K\sigma_Z^2 = N_0 B$, where σ_Z^2 is the noise at each subcarrier, i.e., in DS-CDMA the power from the entire band causes noise to the receiver. The outage probability (73) is then given by

$$q_{DS}(\lambda) = \mathbb{P}\left(\frac{\rho r^{-n}}{K\sigma_Z^2 + \sum_{l \in \Pi_k(\lambda_k)} \rho r_l^{-n}} < \frac{\beta}{K}\right).$$

From the upper bound (75), we see that $c_{DS}(\epsilon) \leq \frac{\epsilon}{\pi}(\kappa')^{-2/n}$, where now $\kappa' = \frac{r^{-n}}{\beta/K} - K\sigma_Z^2/\rho = K\kappa$, which after placing in into (75), concludes the proof. □

Comparing the transmission capacity upper bounds for FH-CDMA (77) and DS-CDMA (78), we observe that the ratio $c_{FH}(\epsilon)/c_{DS}(\epsilon)$ only depends on the spreading factor K and the path-loss parameter n by

$$\frac{c_{FH}(\epsilon)}{c_{DS}(\epsilon)} = K^{1-2/n}.$$

In other words, if $n > 2$, FH-CMDA outperforms DS-CDMA for fixed system conditions by a factor of, e.g. $\sqrt{K}$ in indoor environment with $n = 4$ [78]. For satellite communications and spherical transmission, $n = 2$ and consequently both techniques suffer from the same transmission capacity bounds. Conversely, in some situations such as ultra wideband (UWB) communications, different signals can add constructively at the receiver, leading to a path-loss exponent $n < 2$, for which DS-CDMA is preferred to FH-CDMA. Figure 11 reflects this result, which depicts the transmission capacity for both MATs for nominal SNR = 0 dB, $\beta = -10$ dB, outage probability $\epsilon = 0.1$, and a spreading factor $K = 10$.

4.5.4 Asymptotic Analysis

In the next sections we analyze the behavior of the transmit capacities derived in this Section for asymptotically wideband and high-SNR regimes. We assume that the information signal has a fixed bandwidth of B/K Hz, hence increasing the spreading factors increases the total bandwidth. We now particularize (77) and (78) for the wideband regime. Let the total bandwidth $B \to \infty$. Since the narrow bandwidth required for the information signal is fixed, this is equivalent to consider $K \to \infty$. Under this assumption, we see that the transmit capacity for FH-CDMA (77) increases with a rate which is linear with K. On the other hand, we observe that the transmit capacity of DS-CDMA (78) increases with a sublinear rate of $K^{2/n}$, when $n > 2$, the rate becomes linear for spherical propagation $n = 2$. The explanation for this phenomenon is that when increasing the total bandwidth or spreading factor in DS-CDMA, the total noise power increases linearly with K, whereas the SINR requirement is divided by the same factor K. As a result, the interference caused by the near terminals can be neglected compared to the noise power and, due to path-loss propagation, the outage probability is dominated by the SNR level, which in view of (78) increases the capacity with exponent $2/n$. The effect of the bandwidth, or equivalently spreading factor, can be seen on Figure 12, which depicts the evolution of the transmission

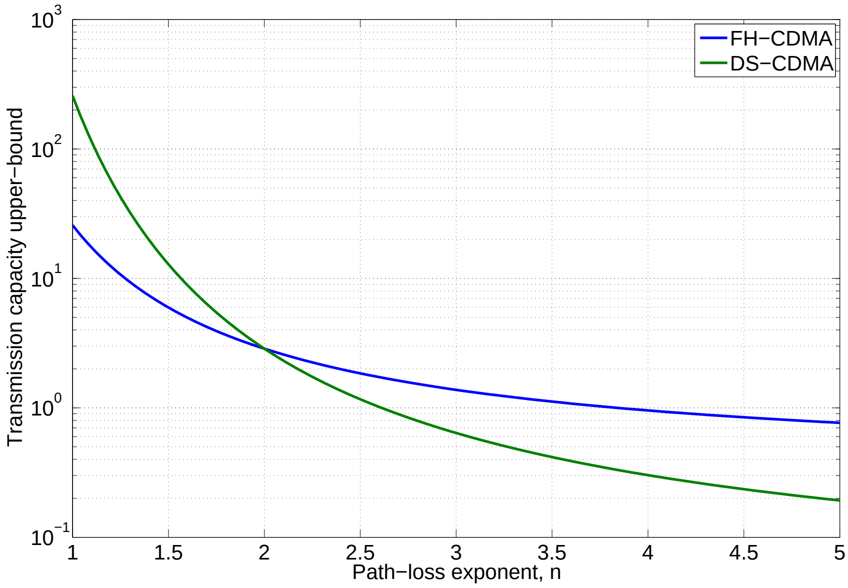

Figure 11: Transmission capacities upper-bounds for FH-CDMA (77) and DS-CDMA (78) versus the path-loss exponent for SNR = 0 dB, $\beta = -10$ dB, $\epsilon = 0.1$ and $K = 10$.

capacities versus the spreading factor K for SNR = 0 dB, SINR threshold $\beta = -10$ dB, outage probability $\epsilon = 0.1$ and path-loss exponent $n = 3$.

Finally, we analyze the asymptotic behaviors with the nominal SNR, defined as SNR $\doteq \rho/\sigma_Z^2$. From (77) and (78), we observe that the parameter κ can be expressed as $\kappa = r^{-n}/\beta - 1/\text{SNR}$. Therefore, for SNR$\rightarrow \infty$, the first term in κ, i.e., r^n/β dominates and the corresponding transmission capacity for FH-CDMA and DS-CDMA saturate at

$$c_{FH}(\epsilon) \rightarrow \frac{\epsilon K}{\pi} \frac{1}{r^2 \beta^{2/n}} \text{ and } c_{FH}(\epsilon) \rightarrow \frac{\epsilon}{\pi} \frac{K^{2/n}}{r^2 \beta^{2/n}},$$

respectively.

In order to illustrate this result, Figure 13 shows the transmission capacity upper-bounds for a spreading factor of $K = 10$, SINR threshold $\beta = 0$ dB, outage probability $\epsilon = 0.1$, and path-loss exponent $n = 3$. The saturation curves for SNR $\rightarrow \infty$ are likewise included. It is important to highlihgt that the transmission capacity is associated to an outage probability on the received SINR being below a required threshold. This means that at the low-SNR regime, the inequality holds for SNR $< \beta$ with an outage probability associated to ϵ. The upper-bound derived in (75), nonetheless, establishes an inequality on the transmission capacity taking into account only the near terminals, which are located geometrically inside a disk of radius $\kappa^{-1/n}$, which must satisfy $\kappa > 0$. This condition reflects that, in general, as $\kappa \doteq r^{-n}/\beta - 1/\text{SNR}$, SNR $> \beta/r^{-n}$, that is, the target SINR scaled by the path-loss propagation. This effect is reflected in Figure 13, as the transmission capacity is valid for the SNR $> \beta/r^{-n}$ region.

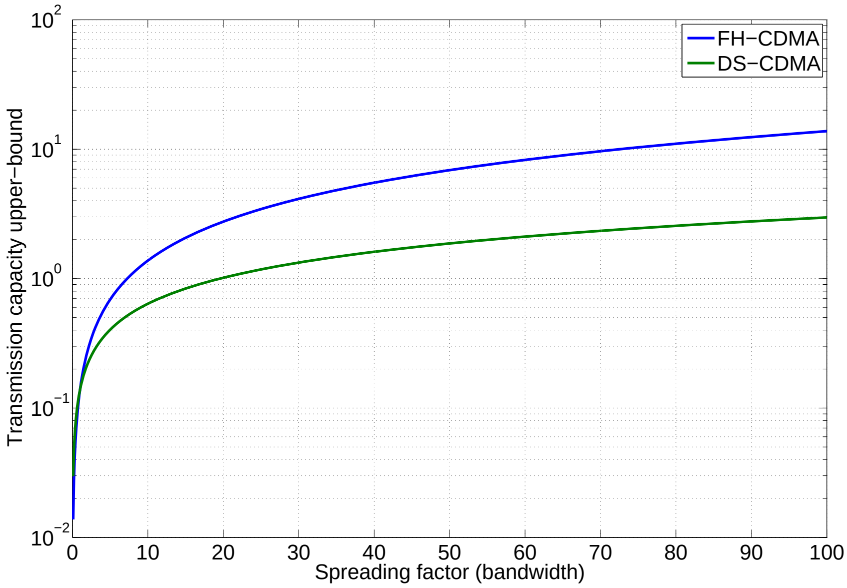

Figure 12: Transmission capacities upper-bounds for FH-CDMA (77) and DS-CDMA (78) versus the spreading factor (bandwidth) for SNR = 0 dB, $\beta = -10$ dB, $\epsilon = 0.1$ and $n = 3$.

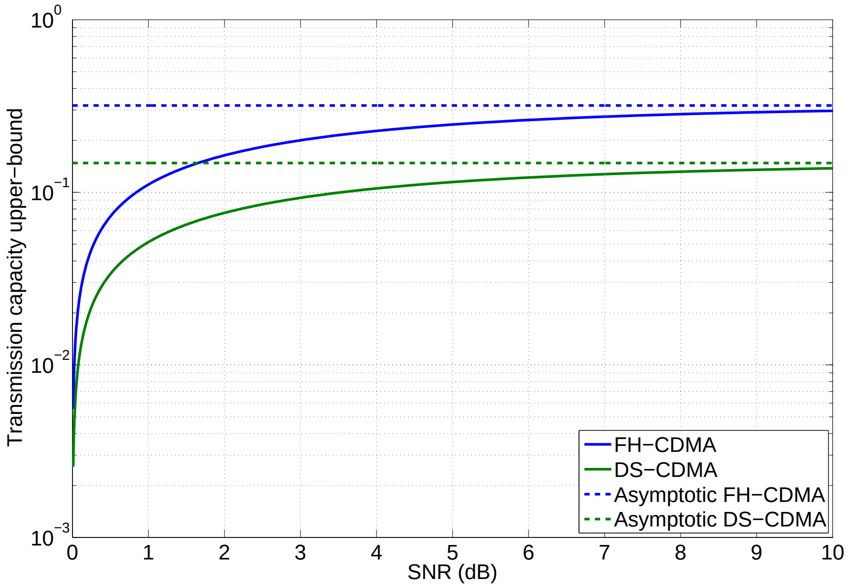

Figure 13: Transmission capacities upper-bounds for FH-CDMA (77) and DS-CDMA (78) versus nominal SNR $K = 10$, $\beta = -10$ dB, $\epsilon = 0.1$ and $n = 3$.

4.5.5 Conclusions

We have shown that TDMA becomes optimal in narrow-band or high-SNR regimes without near-far effect between user terminals; whereas for the wideband or low-SNR regimes, a spread-spectrum strategy such as CDMA should be employed as the noise contribution is dominant. CDMA-based multiple-access strategies tolerate all sources of interference with bounds determined by the spreading factor K, whereas orthogonality in TDMA or FDMA can be compromised due to interferences. CDMA requires that both transmitter and receiver have knowledge of the agreed pseudo-random sequences and the position of the sequence, which is more stringent in DS-CDMA as the synchronization must be established to within a fraction of chirp interval of duration $1/B$, whereas on the other hand in FH-CDMA the chirp interval is the time spent in transmitting the signal in a particular frequency slot of bandwidth B/K, i.e., approximately an interval of duration K/B [13]. Asymptotic performance results illustrate that the transmission capacity for DS-CDMA grows sub-linearly with K while in FH-CDMA it grows linearly for $n > 2$. Hence, increasing the spreading factor in FH-CDMA results in improved interference avoidance, which is preferable to interference suppression such as in DS-CDMA [78].

5 Work Plan

The presentation of this Ph.D. thesis proposal involves the completion of some of the objectives defined in Section 2.2. The most important efforts have been concentrated on acquiring advanced technical background and state-of-the-art signal processing and communication tools for compressed-sampling, spectrum sensing and traffic-capacity theories.

5.1 Lines of Investigation

So far, preliminary results on GLRT-based spectrum sensing for wideband cognitive radio have been achieved, as well as first conclusions within the information-theoretical analysis of the compression matrix and spectral analysis of sparse signals. A first approach for second-order statistics reconstruction from compressed observations based on a correlation-matching approach has also been studied. Furthermore, in the field of dynamic spectrum access, asymptotic results on the traffic capacity theory for wideband regimes have been also achieved.

Therefore, in view of the preliminary results and state-of-the art literature, the following three open investigation lines define the future research for the present Ph.D. thesis:

- Optimal spectrum sensing detection based on compressed-sampling second-order-statistics for wideband cognitive radio (Objectives 1-6).
- Assessment on multiple-access techniques for wideband cognitive radio (Objectives 7-8).
- Underlying wideband cognitive radio communication strategies (Objective 9).

5.2 Methodology

The following tasks, defined for each objective encompassed by the open lines of investigation uncovered above, outline the general means through which the research objectives are achieved.

1. *Optimal spectrum sensing framework for cognitive radio.*

    (a) (Achieved) Derivation of optimal GLRT spectrum sensing detectors for Gaussian noise and signal under the low-SNR regime. The formulation of GLRT detectors include solving convex optimization problems to compute the maximum likelihood (ML) estimates of the statistical parameters of the primary users' signal and noise unknown to the cognitive user, as well as determining the likelihood ratio test employing such estimates.

    (b) (Achieved) Discovering an underlying framework for the GLRT detectors in the frequency-domain based on the second-order statistics of the observations. A comparison between the derived detectors uncovers a general structure for any GLRT detector in terms of a set of parameters that determine the factors that affect the final performance of the detector.

    (c) (Further analysis) Simulation assessments on the performance of the GLRT detectors in realistic scenarios. The metrics employed in the literature are the probability of detection and probability of false alarm restrictions (receiver operating characteristics, ROC) in distinct channel conditions (SNR, primary user correlation level).

2. *Information-theoretical analysis of compressed and nonuniform sampling.*

    (a) (Achieved) Computation of capacity formulas for the compressed-sampling scheme. We identify the compressed-sampling scheme as a communication channel with additive noise before compression, and derive the achievable rates in function of the compression matrix.

    (b) (Further analysis) Particularization of the capacity formulas for wideband regimes. The expressions of the achievable rates are obtained for the low-SNR regime to obtain the asymptotic behavior of the compressed-sampling scheme.

    (c) Comparison of variable and fixed compression matrices. Employing the aforementioned metrics, a critical assessment on the compressed-sampling scheme using fixed or random compression of the observations is required to determine which compression scheme favors the reconstruction of second-order statistics.

3. *Spectral analysis of sparse wideband signals.*

    (a) (Achieved) Derivation of the compressed versions of classical spectral analysis methods such as the Blackman-Tuckey or the Capon level estimators. We recognize that in wideband regimes the activity of the primary users is a sparse signal, and therefore consider a compressed-sampling scheme for spectral analysis.

    (b) Simulation results of sparse wideband spectral analysis in realistic scenarios. This task involves the reconstruction of the primary users' spectrum via Blackman-Tuckey and Capon level estimates in several compression and channel conditions, and determine which compression conditions are good in recovering sparse spectrum signals.

4. *Second-order statistics reconstruction from compressed observations.*

    (a) (Achieved) Compressed correlation-matching for second-order primary users' statistics. Analyzing the preliminary results on the GLRT detectors, we identify that the second-order statistics of the observations become a sufficient statistic for detection. Therefore, a first attempt to recover the sample covariance matrix from the compressed observations is given employing the correlation-matching approach.

(b) Simulation results on the performance of the correlation-matching approach for second-order statistics reconstruction. The performance of the proposed correlation-matching approach is to be determined in terms of mean squared error (MSE) of the reconstructed parameters in realistic scenarios.

5. *Compressed spectrum sensing algorithms for wideband cognitive radio.*

   (a) Combination of the second-order statistics reconstruction from compressed observations with the optimal GLRT spectrum sensing detectors for wideband signals. Once the estimates of the unknown statistical parameters of the primary users' signal and noise are obtain via compressed-sampling reconstruction, the optimal likelihood ratio test with compressed observations is obtained.

   (b) Unified framework of compressed GLRT spectrum sensing detectors. This task involves the further analysis on the frequency-domain parameters that determine the general structure of the detection when using compressed observations.

   (c) Final performance of compressed spectrum sensing algorithms. The ROC of the proposed compressed detectors is to be obtained in several primary users' signal, channel and compression conditions, and a comparison to other state-of-the art compressed-sampling spectrum sensing approaches is required.

6. *Cooperative spectrum sensing of sparse signals.*

   (a) Derivation of optimal GLRT spectrum sensing detectors with neighbors side information. The information on the detection or ML estimation performed by nodes in the vicinity can be incorporated in the GLRT detector as prior information. Then, each node performs a maximum a posteriori (MAP) estimation of the statistics parameters for detection based on the prior information passed by the adjacent nodes and the compressed local observations.

   (b) Design of message-passing strategies for side information communication among nodes in the area. This task involves determining the most relevant information that affects the individual spectrum sensing detection, and how this information is sent to the other nodes, by establishing a trade-off between performance and network overhead.

   (c) Performance of cooperative spectrum sensing. In this task, the ROC of the proposed cooperative detectors has to be determined to uncover which parameters of the compression and cooperation schemes play a fundamental role in the detection of primary user activity in wideband regimes.

7. *Traffic capacity asymptotic performance in wideband cognitive radios.*

   (a) (Achieved) Background on the novel formulation of traffic capacity. A detailed view on the premises, derivations, and conclusions of the traffic capacity theory is done. This study concludes that main restrictions and mathematical tools required to apply this field to cognitive radio.

(b) (Further analysis) Particularization of traffic capacity formulas for wideband regimes in practical dynamic access techniques. The asymptotic behavior of the frequency-hopping and direct-sequence CDMA traffic capacities is studied in this task.

(c) Traffic capacity formulation of cognitive radio networks with primary user interference restrictions. In underlying and interweave cognitive radios, the secondary user network is conditioned by the primary user activity, i.e., the spectrum sensing information and the interference caused to the primary users. Therefore, a traffic capacity analysis taking into consideration these factors is done in this task.

8. *Multiple-access techniques for cognitive radio.*

   (a) (Further analysis) Classical and traffic capacity assessment on multiple-access techniques. A fair comparison between multiple-access techniques in terms of classical multiple-access capacity and the novel traffic capacity perspectives is required to determine which techniques are suited for cognitive radio networks.

   (b) Design of multiple-access techniques for wideband cognitive radio. Recognizing the sparse and interference-limited nature of cognitive radio, the design of efficient multiple-access techniques of secondary users is done by optimizing the capacity of the secondary network by making use of the spectrum sensing information on the activity of the primary users.

9. *Communication approaches for underlying cognitive radios.*

   (a) Final combination of compressed-sampling, spectrum sensing and dynamic access for wideband cognitive radio. The precedent tasks are designed to uncover which are the optimal compression, sensing and access conditions in realistic cognitive radio scenarios where the primary users' activity is sparse in wideband regimes. In this task, we will combine the three fields to check the final performance of the physical and access layers in cognitive radio.

   (b) Alternative methods for underlying cognitive radio communication. Whereas task 9a focuses on cross-layer optimization, communication approaches can be designed directly from the compressed local observations to perform opportunistic communication in cognitive radio networks, e.g., by transmitting in such a way that the interference caused to the primary systems is small without performing spectrum sensing.

In view of the preliminary work achieved, the aforementioned pending and future tasks, and taking into consideration that this Ph.D. thesis has started in September 2009, a tentative date for the Ph.D. thesis dissertation is May 2012.

Appendices

A Remark for the Proof of Proposition 2

Recalling the expression of the complex multivariate Gaussian distribution, the exact LRT for the detection problem (16) when the noise and signal covariance matrices are known can be written as

$$L(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) = \frac{\det(\mathbf{C}_W)^M}{\det(\mathbf{C}_W + \mathbf{C}_S)^M} \cdot \frac{\exp \operatorname{tr}(-\mathbf{X}^H(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{X})}{\exp \operatorname{tr}(-\mathbf{X}^H \mathbf{C}_W^{-1}\mathbf{X})} \tag{79}$$

Consider the matrix inversion lemma $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$, for invertible matrices. When applied to $(\mathbf{C}_S + \mathbf{C}_W)^{-1}$, the inversion lemma results in $(\mathbf{C}_S + \mathbf{C}_W)^{-1} = \mathbf{C}_W^{-1} - \mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_W^{-1}\mathbf{C}_S + \mathbf{I})^{-1}\mathbf{C}_W^{-1}$. By left-multiplying the inside of the inverse term $(\mathbf{C}_W^{-1}\mathbf{C}_S + \mathbf{I})^{-1}$ by $\mathbf{C}_W$, it follows that $(\mathbf{C}_S + \mathbf{C}_W)^{-1} = \mathbf{C}_W^{-1} - \mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}$. Combining this result with the second fraction in (79), the LRT is then given by

$$L(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) = \frac{\det(\mathbf{C}_W)^M}{\det(\mathbf{C}_W + \mathbf{C}_S)^M} \cdot \exp \operatorname{tr}(\mathbf{X}^H \mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{X}).$$

Finally, the log-LRT, defined as the logarithm of the LRT, is given by

$$\begin{aligned} l(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) &\doteq \ln L(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) \\ &= M \ln \det(\mathbf{C}_W) - M \ln \det(\mathbf{C}_S + \mathbf{C}_W) + \operatorname{tr}\left(\mathbf{X}^H \mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{X}\right), \end{aligned} \tag{80}$$

whereas the equivalent test $T(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W)$ in (20) is obtained retaining only the data-dependent terms, i.e., the term $\operatorname{tr}\left(\mathbf{X}^H \mathbf{C}_W^{-1}\mathbf{C}_S(\mathbf{C}_S + \mathbf{C}_W)^{-1}\mathbf{X}\right)$. Since the logarithm is a monotonically increasing function and the remaining term is positive, the final test decides for $\mathcal{H}_1$ is $T(\mathbf{X}|\mathbf{C}_S, \mathbf{C}_W) \geq \gamma$. This is an important remark, as in detection simpler expressions are preferred. Finally, moving $\mathbf{X}^H$ into the right-hand side of the trace, we equivalently obtain (20) up to a non-negative scaling factor.

B Proof of Theorem 1

Departing from the expression of the optimal test when the noise and signal covariance matrices are known (20), we develop the $(\mathbf{C}_S + \mathbf{C}_W)^{-1}$ term in the following manner. We take $\mathbf{C}_W$ as a common left-factor, i.e., $[\mathbf{C}_W(\mathbf{C}_W^{-1}\mathbf{C}_S + \mathbf{I})]^{-1}$, which after making use of the low-SNR Approximation (81) in Appendix C it becomes $\mathbf{C}_W^{-1}$, which concludes the proof.

C Low-SNR Approximation

Let $\mathbf{C}_S$ and $\mathbf{C}_W$ be the covariance matrices of the signal and the noise, respectively. At low-SNR, asymptotically

$$\mathbf{C}_W^{-1}\mathbf{C}_S + \mathbf{I} \to \mathbf{I}. \tag{81}$$

Proof. Let us express the signal and noise covariance matrices in the following form,

$$\begin{aligned}\mathbf{C}_S &\doteq \sigma_S^2 \mathbf{\Sigma}_S,\\ \mathbf{C}_W &\doteq \sigma_W^2 \mathbf{\Sigma}_W,\end{aligned} \tag{82}$$

where $\sigma_S^2, \sigma_W^2 \geq 0$ are the variances of the signal and noise, and $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_W$ have ones at the main diagonal, i.e., $[\mathbf{\Sigma}_S]_{nn} = [\mathbf{\Sigma}_W]_{nn} = 1$ for $1 \leq n \leq N$. Then, the SNR is reflected in the product $\mathbf{C}_W^{-1}\mathbf{C}_S = \sigma_S^2/\sigma_W^2 \mathbf{\Sigma}_W^{-1}\mathbf{\Sigma}_S$, which at the low-SNR regime, approaches $\sigma_S^2/\sigma_W^2 \to 0$. □

D Proof of Theorem 2

The ML estimate of $\mathbf{C}_S$ is given, according to the GLRT estimates (18) by the maximization problem

$$\begin{aligned}\hat{\mathbf{C}}_S = \arg\max\nolimits_{\mathbf{C}_S} p(\mathbf{X}|\mathcal{H}_1, \mathbf{C}_S, \mathbf{C}_W)\\ \text{s.t.} \mathbf{C}_S \succeq \mathbf{0}.\end{aligned} \tag{83}$$

The solution to the former problem is obtained by solving the Karush-Kuhn-Tucker (KKT) conditions [107] given by

- The stationarity condition $\nabla\mathcal{L}(\hat{\mathbf{C}}_S) = 0$,
- the primal feasibility $\lambda_n(\hat{\mathbf{C}}_S) \geq 0$ for $1 \leq n \leq N$,
- the dual feasibility $\mu_n \leq 0$ for $1 \leq n \leq N$, and
- the complementary slackness $\mu_n\lambda_n(\hat{\mathbf{C}}_S) = 0$ for $1 \leq n \leq N$,

where $\mathcal{L}(\mathbf{C}_S)$ is the Lagrangian $\mathcal{L}(\mathbf{C}_S) = \ln p(\mathbf{X}|\mathcal{H}_1, \mathbf{C}_S, \mathbf{C}_W) + \sum_{n=1}^{M} \mu_n\lambda_n(\mathbf{C}_S)$, and μ_n are the Lagrange multipliers for the inequality constraint.

First, from the stationarity condition we obtain that the eigenvalue decomposition of $\hat{\mathbf{C}}_S$ must accomplish $\hat{\mathbf{C}}_S = \hat{\mathbf{U}}_S\hat{\mathbf{\Lambda}}_S\hat{\mathbf{U}}_S^H = \mathbf{A} - \mathrm{diag}\,(\mu_1, \ldots, \mu_N)$, where $\mathbf{A} \doteq \mathbf{R}_X - \mathbf{C}_W$. Applying the primal feasibility and taking into account the dual feasibility, we see that the Lagrange multipliers are non-zero for those eigenvalues of $\mathbf{A}$ that are negative, that is, $\mu_n = \min(\lambda_n(\mathbf{A}), \mathbf{0})$, for which the complementary slackness is accomplished as in such a case $\lambda_n(\hat{\mathbf{C}}_S) = 0$, otherwise $\mu_n = 0$. Finally, we then obtain that $\hat{\mathbf{U}}_S = \mathbf{U}_A$ and $\hat{\mathbf{\Lambda}}_S = \mathbf{\Lambda}_A^+$, where $\mathbf{\Lambda}_A^+$ denotes diagonal-wise $\max(\lambda_n(\mathbf{A}), 0)$, which concludes the first part of the proof.

With the $\hat{\mathbf{C}}_S$ estimate, the test (79) becomes

$$L(\mathbf{X}|\mathbf{C}_W) = \frac{\det(\mathbf{C}_W)^M}{\det(\hat{\mathbf{C}}_S + \mathbf{C}_W)^M} \frac{\exp\mathrm{tr}(-\mathbf{X}(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{X})}{\exp\mathrm{tr}(-\mathbf{X}\mathbf{C}_W^{-1}\mathbf{X})},$$

which can be further processed as in Appendix A to obtain the log-GLRT (80) with placing the estimate of $\mathbf{C}_S$ instead of its known value, i.e.,

$$l(\mathbf{X}|\mathbf{C}_W) \doteq \ln L(\mathbf{X}|\mathbf{C}_W) = M\ln\det(\mathbf{C}_W) - M\ln\det(\hat{\mathbf{C}}_S + \mathbf{C}_W) + \mathrm{tr}\left(\mathbf{X}^H\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{X}\right).$$

Now, the only data independent term is $M \ln \det \mathbf{C}_W$, which after being removed from $l(\mathbf{X}|\mathbf{C}_W)$, we obtain (22) up to a nonnegative scaling factor M. Test (22) allows further simplification for the following two particular cases.

On the one hand, consider the asymptotic PDF of zero-mean wide sense stationary Gaussian random processes. Clearly, $\mathbf{R}_X$ obeys a symmetric Toeplitz matrix and, as such, it is known that for $N \to \infty$ (i.e., large data records), the eigenvalues $\lambda_n(\mathbf{R}_X)$ and eigenvectors $\mathbf{u}_n(\mathbf{R}_X)$ are easily found by [94]

$$\begin{gathered}\lambda_n(\mathbf{R}_X) = S_X(f_n), \\ \mathbf{u}_n = \tfrac{1}{\sqrt{N}}\left[1 \ \exp(\jmath 2\pi f_n) \ \ldots \ \exp(\jmath 2\pi (N-1) f_n)\right]^T\end{gathered} \tag{84}$$

where $S_X(f)$ denotes the power spectral density of the observations and $f_n \doteq n/N$. The same result can be applied to the noise process. As a result, $\mathbf{U}_X = \mathbf{U}_W$ and, by construction, also equal to $\hat{\mathbf{U}}_S$ so that we obtain that the ML estimate for the signal covariance matrix obeys

$$\hat{\mathbf{C}}_S = \mathbf{U}\left(\mathbf{\Lambda}_X - \mathbf{\Lambda}_W\right)^+ \mathbf{U}^H, \tag{85}$$

where $\mathbf{U} = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_N]$. By taking the logarithm to $L(\mathbf{X}|\mathbf{C}_W)$ we then obtain the following four terms

$$\begin{aligned}
M \ln \det(\mathbf{C}_W) &= M \sum_{n=1}^{N} \ln \lambda_n(\mathbf{C}_W) \\
-M \ln \det(\hat{\mathbf{C}}_S + \mathbf{C}_W) &= -M \sum_{n=1}^{N} \ln \max\left(\lambda_n(\mathbf{R}_X), \lambda_n(\mathbf{C}_W)\right) \\
-\mathrm{tr}(\mathbf{X}^H(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{X}) &= -M \sum_{n=1}^{N} \min\left(1, \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}\right) \\
\mathrm{tr}(\mathbf{X}^H \hat{\mathbf{C}}_W^{-1} \mathbf{X}) &= M \sum_{n=1}^{N} \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}
\end{aligned}$$

By adding the fourth terms up, the equivalent log-GLRT is given by

$$l(\mathbf{X}|\mathbf{C}_W) = M \sum_{n=1}^{N} \left(\ln \frac{\lambda_n(\mathbf{C}_W)}{\max\left(\lambda_n(\mathbf{R}_X), \lambda_n(\mathbf{C}_W)\right)} - \min\left(1, \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}\right)\right) + M \sum_{n=1}^{N} \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)}$$

We observe from the former expression that the $\max$ and $\min$ operators keep only a few data-dependent terms, leading to

$$T(\mathbf{X}|\mathbf{C}_W) = \sum_{n \in \mathcal{A}_+} \ln \frac{\lambda_n(\mathbf{C}_W)}{\lambda_n(\mathbf{R}_X)} - \sum_{n \in \mathcal{A}_+} \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)},$$

which after a sign multiplication, concludes the second part of the proof. Here, the set $\mathcal{A}_+$ has been defined as the set of indexes for which the eigenvalues of the observations are larger than the eigenvalues of the noise, i.e., $\mathcal{A}_+ \doteq \{n \,:\, \lambda_n(\mathbf{R}_X) > \lambda_n(\mathbf{C}_W)\}$

Finally, under the white-noise assumption, we observe that matrix $\mathbf{A}$ takes the eigenvectors of $\mathbf{R}_X$, as $\mathbf{C}_S$ does so. In such a case, we similarly obtain that the ML estimate for the signal covariance matrix obeys $\hat{\mathbf{C}}_S = \mathbf{U}_X\left(\mathbf{\Lambda}_X - \mathbf{\Lambda}_W\right)^+ \mathbf{U}_X^H$, and after similar mathematical manipulations, $T(\mathbf{X}|\mathbf{C}_W)$ becomes

(24), that is, with $\lambda_n(\mathbf{C}_W) = \sigma_W^2$ for $1 \leq n \leq N$.

E Proof of Theorem 3

We consider the low-SNR approximation for the optimal test when the noise signal covariance matrix is unknown, given by (22), which we recall here

$$T_2(\mathbf{X}|\mathbf{C}_W) = \mathrm{tr}\left(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{R}_X\right) - \ln\det(\hat{\mathbf{C}}_S + \mathbf{C}_W).$$

For the first part of the test, following analogous developments as for the estimator-correlator (21), we can state that

$$\mathrm{tr}\left(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S(\hat{\mathbf{C}}_S + \mathbf{C}_W)^{-1}\mathbf{R}_X\right) \to \mathrm{tr}\left(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S\mathbf{C}_W^{-1}\mathbf{R}_X\right), \tag{86}$$

whereas for the second part of the test, we make us of the following approximation of the logarithm of the determinant for small-normed matrices. Specifically, the first-order approximation $\ln\det(\mathbf{I}+\epsilon\mathbf{A}) \to \epsilon\mathrm{tr}(\mathbf{A})$, when $\epsilon \to 0$ can be cast with $\ln\det(\mathbf{I} + \mathbf{C}_W^{-1}\hat{\mathbf{C}}_S)$ when using the low-SNR approximation given in Appendix C, with $\epsilon = \sigma_S^2/\sigma_W^2$. Therefore, the second part of the test can be further developed as

$$\ln\det(\hat{\mathbf{C}}_S + \mathbf{C}_W) = \ln\det(\mathbf{I} + \mathbf{C}_W^{-1}\hat{\mathbf{C}}_S) + \ln\det\mathbf{C}_W \to \mathrm{tr}(\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S) + \ln\det\mathbf{C}_W. \tag{87}$$

By neglecting the data independent term $\ln\det\mathbf{C}_W$ and joining (86) with (87), we proof (22).

We now show the low-SNR approximation for the white noise assumption, which is also valid and equivalent to large data records. At the low-SNR regime, we can assume that the ratio $\lambda_n(\mathbf{R}_X)/\lambda_n(\mathbf{C}_W)$ in (26) approaches to one, as $\lambda_n(\mathbf{R}_X)$ contains the power of both noise and signal.

Consider now the Taylor-decomposition of the natural logarithm function, i.e., $\ln(1+x) = x - x^2/2 + \mathcal{O}(x^3)$ for $x \to 0$. When applied to $1 + x = \lambda_n(\mathbf{R}_X)/\lambda_n(\mathbf{C}_W)$, we note that at the low-SNR regime, by neglecting the higher order tems, we get

$$\ln\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)} = \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)} - 1 - \frac{1}{2}\left(\frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_W)} - 1\right)^2$$

.

After some mathematical manipulations and omitting data-independent terms, the test becomes (26), concluding the proof.

F Proof of Theorem 4

The ML estimates of the noise covariance matrix for the cognitive radio problem (16) when the signal covariance matrix is known are given by the maximization problems (18)

$$\hat{\mathbf{C}}_{W_0} = \arg\max_{\mathbf{C}_W} p(\mathbf{X}|\mathcal{H}_0, \mathbf{C}_W)$$
$$\hat{\mathbf{C}}_{W_1} = \arg\max_{\mathbf{C}_W} p(\mathbf{X}|\mathcal{H}_1, \mathbf{C}_S, \mathbf{C}_W),$$

both subject to $\mathbf{C}_W \succeq \mathbf{0}$. On the one hand, the ML estimate under $\mathcal{H}_0$ is given by [44] $\hat{\mathbf{C}}_{W_0} = \mathbf{R}_X$. On the other hand, the ML estimate under $\mathcal{H}_0$ is analogous to problem (83), i.e., the ML estimate of $\mathbf{C}_S$

when the noise variance matrix is known. Hence, $\hat{\mathbf{C}}_{W_1}$ is given by $\mathbf{U}_B\mathbf{\Lambda}_B^{+}\mathbf{U}_B^H$, where $\mathbf{U}_B\mathbf{\Lambda}_B\mathbf{U}_B^H$ is the eigenvalue decomposition of the correlation matrix $\mathbf{B} \doteq \mathbf{R}_X - \mathbf{C}_S$ and the superscript in $\mathbf{\Lambda}_B^{+}$ denotes the diagonal-wise operation $\max(\lambda_n(\mathbf{B}), 0)$, concluding the first part of the proof.

With the $\hat{\mathbf{C}}_{W_0}$ and $\hat{\mathbf{C}}_{W_1}$ estimates, the test (79) becomes

$$L(\mathbf{X}|\mathbf{C}_S) = \frac{\det(\hat{\mathbf{C}}_{W_0})^M}{\det(\mathbf{C}_S + \hat{\mathbf{C}}_{W_1})^M} \cdot \frac{\exp\operatorname{tr}(-\mathbf{X}^H(\mathbf{C}_S + \hat{\mathbf{C}}_{W_1})^{-1}\mathbf{X})}{\exp\operatorname{tr}(-\mathbf{X}^H\hat{\mathbf{C}}_{W_0}^{-1}\mathbf{X})},$$

which after some manipulations, the log-GLRT $l(\mathbf{X}|\mathbf{C}_S)$ derives to (27). We now focus on two particular assumptions. First, we consider the asymptotic case for large data records. The PDF of the zero-mean wide sense stationary Gaussian random processes corresponding to the observations and the noise. Clearly, both $\mathbf{R}_X$ and $\mathbf{C}_W$ will obey a Toeplitz matrix whose eigenvectors, for $N \to \infty$, are given by (84). As a result, matrix $\mathbf{B}$ will do so. Then, by taking the logarithm of the GLRT, we obtain the following four terms:

$$\begin{aligned}
M\ln\det(\hat{\mathbf{C}}_{W_0}) &= M\sum_{n=1}^{N}\ln\lambda_n(\mathbf{R}_X) \\
-M\ln\det(\mathbf{C}_S + \hat{\mathbf{C}}_{W_1}) &= -M\sum_{n=1}^{N}\ln\max\left(\lambda_n(\mathbf{R}_X), \lambda_n(\mathbf{C}_S)\right) \\
-\operatorname{tr}(\mathbf{X}^H(\mathbf{C}_S + \hat{\mathbf{C}}_{W_1})^{-1}\mathbf{X}) &= -M\sum_{n=1}^{N}\min\left(1, \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_S)}\right) \\
\operatorname{tr}(\mathbf{X}^H\hat{\mathbf{C}}_{W_0}^{-1}\mathbf{X}) &= MN.
\end{aligned}$$

By adding them up, we obtain equivalently that the log-GLRT is given by

$$l(\mathbf{X}|\mathbf{C}_S) = M\sum_{n=1}^{N}\left(\ln\frac{\lambda_n(\mathbf{R}_X)}{\max\left(\lambda_n(\mathbf{R}_X), \lambda_n(\mathbf{C}_S)\right)} - \min\left(1, \frac{\lambda_n(\mathbf{R}_X)}{\lambda_n(\mathbf{C}_S)}\right)\right).$$

We see from the former expression that only the eigenvalues accomplishing $\lambda_n(\mathbf{R}_X) > \lambda_n(\mathbf{C}_S)$ are the remaining data-dependent terms, leading to (28) and concluding the second part of the proof.

G Proof of Theorem 6

Consider the optimal test for unknown white noise and known signal covariance matrix, given by

$$L(\mathbf{X}|\mathbf{C}_S) = \frac{\det(\hat{\sigma}_{W_0}^2\mathbf{I})^M}{\det(\hat{\sigma}_{W_1}^2\mathbf{I} + \mathbf{C}_S)^M} \cdot \frac{\exp\operatorname{tr}\left(-\mathbf{X}^H(\hat{\sigma}_{W_1}^2\mathbf{I} + \mathbf{C}_S)^{-1}\mathbf{X}\right)}{\exp\operatorname{tr}\left(-\mathbf{X}^H(\hat{\sigma}_{W_0}^2\mathbf{I})^{-1}\mathbf{X}\right)}. \tag{88}$$

We now make use of the following approximation. At the low-SNR regime, consider the signal covariance matrix decomposition (82), with $\mathbf{\Sigma}_W$ the identity matrix. Then, the addition $\hat{\sigma}_{W_1}^2\mathbf{I} + \mathbf{C}_S$ can be written as

$$\hat{\sigma}_{W_1}^2\mathbf{I} + \mathbf{C}_S = \hat{\sigma}_{W_1}^2\mathbf{I} + \sigma_S^2\mathbf{\Sigma}_S = \hat{\sigma}_{W_1}^2(\mathbf{I} + \sigma_S^2/\hat{\sigma}_{W_1}^2\mathbf{\Sigma}_S) \to \hat{\sigma}_{W_1}^2\mathbf{I},$$

where we made use of Approximation (81). After taking the logarithm to (88) we get the following terms

$$\begin{aligned} M \ln \det(\hat{\sigma}^2_{W_0}\mathbf{I}) &\rightarrow MN \ln \hat{\sigma}^2_{W_0} \\ -M \ln \det(\hat{\sigma}^2_{W_1}\mathbf{I} + \mathbf{C}_S) &\rightarrow -MN \ln \hat{\sigma}^2_{W_1} \\ -\mathrm{tr}\left(\mathbf{X}^H(\hat{\sigma}^2_{W_1}\mathbf{I} + \mathbf{C}_S)^{-1}\mathbf{X}\right) &\rightarrow -MN\hat{\sigma}^{-2}_{W_1}\hat{\sigma}^2_{W_0} \\ \mathrm{tr}\left(\mathbf{X}^H(\hat{\sigma}^2_{W_0}\mathbf{I})^{-1}\mathbf{X}\right) &= MN. \end{aligned}$$

Finally, neglecting the data-independent terms in the test, we obtain (31), which concludes the proof.

H Remarks on the Proof of Proposition 3

The ML estimates of $\mathbf{C}_S$ and σ^2_W are given, according to the maximization problem (18), i.e.,

$$\begin{aligned} \hat{\sigma}^2_{W_0} &= \arg\max_{\sigma^2_W} p(\mathbf{X}|\mathcal{H}_0, \sigma^2) \\ \hat{\mathbf{C}}_S + \hat{\sigma}^2_{W_1}\mathbf{I} &= \arg\max_{\mathbf{C}_S + \sigma^2_W\mathbf{I}} p(\mathbf{X}|\mathcal{H}_1, \mathbf{C}_S, \sigma^2_W), \end{aligned}$$

where both estimators have the constraint $\sigma^2_W \geq 0$ and $\mathbf{C}_S + \sigma^2_W\mathbf{I} \succeq \mathbf{0}$, respectively. These are classical estimation problems, whose solution is given by [88] $\hat{\sigma}^2_{W_1} = \mathrm{tr}(\mathbf{R}_X)/M$ and $\hat{\mathbf{C}}_S + \hat{\sigma}^2_{W_1}\mathbf{I} = \mathbf{R}_X$.

As a result, the GLRT (17) becomes

$$L(\mathbf{X}) = \frac{\det(\hat{\sigma}^2_{W_0}\mathbf{I})^M}{\det(\hat{\mathbf{C}}_S + \hat{\sigma}^2_{W_1}\mathbf{I})^M} \cdot \frac{\exp \mathrm{tr}(-\mathbf{X}^H(\hat{\mathbf{C}}_S + \hat{\sigma}^2_{W_1}\mathbf{I})^{-1}\mathbf{X})}{\exp \mathrm{tr}(-\mathbf{X}^H\hat{\sigma}^{-2}_{W_0}\mathbf{X})}. \tag{89}$$

Given the expressions of the ML estimates, both terms inside the trace operators become data-independent constant and can be removed from the test. By taking the logarithm and substituting the expressions of the ML estimates, the test is then given by

$$T(\mathbf{X}) = M \ln \det(\hat{\sigma}^2_{W_0}\mathbf{I}) - M \ln \det(\hat{\mathbf{C}}_S + \hat{\sigma}^2_{W_1}\mathbf{I}) = MN \ln \mathrm{tr}\left(\frac{\mathbf{R}_X}{N}\right) - M \ln \det\left(\mathbf{R}_X\right).$$

By dividing the whole expression by NM, we get

$$\begin{aligned} T(\mathbf{X}) &\equiv \ln \mathrm{tr}\left(\frac{\mathbf{R}_X}{N}\right) - \ln \det\left(\mathbf{R}_X\right)^{1/N} \\ &\equiv \ln \frac{\frac{1}{N}\mathrm{tr}(\mathbf{R}_X)}{\det(\mathbf{R}_X)^{1/N}}, \end{aligned}$$

which after removing the logarithm, concludes the proof.

I Proof of Theorem 8

Under the white noise assumption $\mathbf{C}_W = \sigma^2_W\mathbf{I}$, we consider the following structure for the signal covariance matrix $\mathbf{C}_S = \sigma^2_S\mathbf{C}_0$, with $\mathbf{C}_0$ normalized as $\mathrm{tr}(\mathbf{C}_0) = N$. The correlation-matching criterion is given by the optimization problem

$$\hat{\sigma}^2_{W_1}, \hat{\sigma}^2_S = \arg\min_{\sigma^2_W, \sigma^2_S} \|\mathbf{R}_X - \sigma^2_S \mathbf{C}_0 - \sigma^2_W \mathbf{I}\|^2_F.$$

By taking the derivative with respect to σ^2_W and $\sigma^2 S$, we obtain the system of equations

$$\begin{aligned} \mathrm{tr}(\mathbf{R}_X) - N\sigma^2_W - N\sigma^2_S &= 0 \\ \mathrm{tr}(\mathbf{C}_0\mathbf{R}_X) - N\sigma^2_W - \sigma^2_S - \mathrm{tr}(\mathbf{C}_0^2) &= 0, \end{aligned}$$

whose solution is given by (34)–(35), concluding the first part of the proof.

Under $\mathcal{H}_0$, we have similarly that the correlation-matching estimate of the noise variance is given by the optimization problem

$$\hat{\sigma}^2_{W_0} = \arg\min_{\sigma^2_W} \|\mathbf{R}_X - \sigma^2_W \mathbf{I}\|^2_F,$$

which after taking the derivative and equal it to zero, we obtain $\hat{\sigma}^2_{W_0} = \frac{1}{N}\mathrm{tr}(\mathbf{R}_X)$, the same as the ML estimate [88].

J Proof of Theorem 7

Consider the optimal GLRT (32) when both signal and noise covariance matrices are unknown, which we write it in the logarithm version. We now make the following approximation

$$\ln \frac{\frac{1}{N}\mathrm{tr}(\mathbf{R}_X)}{\det(\mathbf{R}_X)^{1/N}} = \ln \frac{\frac{1}{N}\sum_{n=1}^N \lambda_n(\mathbf{R}_X)}{\left(\prod_{n=1}^N \lambda_n(\mathbf{R}_X)\right)^{1/N}} \approx \frac{\sqrt{\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X)}}{\frac{1}{N}\sum_{n=1}^N \lambda_n(\mathbf{R}_X)}.$$

We define p as the average power estimate, $p \doteq \frac{1}{N}\sum_{n=1}^N \lambda_n(\mathbf{R}_X)$. Then,

$$\begin{aligned} \frac{\sqrt{\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X)}}{\frac{1}{N}\sum_{n=1}^N \lambda_n(\mathbf{R}_X)} &= \frac{\sqrt{\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X)}}{p} \\ &= \frac{1}{2}\frac{\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X)}{p^2} \\ &= \frac{1}{2}\ln \frac{p^2 + \left(\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X) - p^2\right)}{p^2} \\ &= \frac{1}{2}\ln\left(1 + \frac{\frac{1}{N}\sum_{n=1}^N \lambda_n^2(\mathbf{R}_X) - p^2}{p^2}\right) \\ &\approx \frac{1}{2p^2 N}\sum_{n=1}^N (\lambda_n(\mathbf{R}_X) - p)^2, \end{aligned}$$

which concludes the proof.

K Proof of Theorem 9

According to the GLRT, the ML estimate of $\boldsymbol{\sigma}^2$ is the solution to the constrained maximization problem

$$\begin{array}{rcl} \hat{\boldsymbol{\sigma}}^2 & = & \arg\max_{\boldsymbol{\sigma}^2} p(\mathbf{X}|\mathbf{C}_W, \mathbf{C}_0, \boldsymbol{\sigma}^2) \\ & & \text{s.t.} \quad \boldsymbol{\sigma}^2 \succeq \mathbf{0}. \end{array} . \tag{90}$$

Let us first consider the log-likelihood function to be maximized, that is,

$$\ln p(\mathbf{X}|\boldsymbol{\sigma}^2) = -MN \ln \pi - M \ln\det(\mathbf{C}_W + \mathbf{C}_S) - M\mathrm{tr}\left((\mathbf{C}_W + \mathbf{C}_S)^{-1}\mathbf{R}_X\right),$$

where $\mathbf{C}_S = \sum_{l=1}^{K} \sigma_l^2 \mathbf{C}_l$. We take into account then the following derivative properties:

$$\begin{array}{rcl} \dfrac{\partial}{\partial x}\mathbf{A}^{-1} & = & -\mathbf{A}^{-1}\dfrac{\partial}{\partial x}\mathbf{A}\mathbf{A}^{-1}, \\ \dfrac{\partial}{\partial x}\ln\det(\mathbf{A}) & = & \mathrm{tr}\left(\mathbf{A}^{-1}\dfrac{\partial}{\partial x}\mathbf{A}\right). \end{array}$$

When applied to the log-likelihood function for an arbitrary σ_k^2, the terms become

$$\begin{array}{rcl} \dfrac{\partial}{\partial \sigma_k^2}(\mathbf{C}_W + \mathbf{C}_S)^{-1} & = & -(\mathbf{C}_W + \mathbf{C}_S)^{-1}\mathbf{C}_k(\mathbf{C}_W + \mathbf{C}_S)^{-1} \\ \dfrac{\partial}{\partial \sigma_k^2}\ln\det(\mathbf{C}_W + \mathbf{C}_S) & = & \mathrm{tr}\left((\mathbf{C}_W + \mathbf{C}_S)^{-1}\mathbf{C}_k\right), \end{array}$$

which after some mathematical manipulations leads to (43), concluding the first part of the proof.

Solving (43) might be difficult due to the inversion of the terms. We now consider the first order Taylor decomposition of the following function $f(\boldsymbol{\sigma}^2) = (\mathbf{I} + \sum_{l=1}^{K} \sigma_l^2/\sigma_W^2 \boldsymbol{\Sigma}_W^{-1}\mathbf{C}_k)^{-1}$ for $\boldsymbol{\sigma}^2/\sigma_W^2 \to \mathbf{0}$. That is,

$$\left[\frac{\partial}{\partial \boldsymbol{\sigma}^2} f(\boldsymbol{\sigma}^2)\right]_k = -\left(\mathbf{I} + \sum_{l=1}^{K} \sigma_l^2/\sigma_W^2 \boldsymbol{\Sigma}_W^{-1}\mathbf{C}_k\right)^{-1} \mathbf{C}_k \left(\mathbf{I} + \sum_{l=1}^{K} \sigma_l^2/\sigma_W^2 \boldsymbol{\Sigma}_W^{-1}\mathbf{C}_k\right)^{-1},$$

and hence, $f(\boldsymbol{\sigma}^2) \approx \mathbf{I} - \sum_{l=1}^{K} \sigma_l^2/\sigma_W^2 \boldsymbol{\Sigma}_W^{-1}\mathbf{C}_k$. Applying this result to the left-hand term of equation (43), we get

$$\mathrm{tr}\left(\left(\mathbf{C}_W + \sum_{l=1}^{K} \sigma_l^2 \mathbf{C}_l\right)^{-1} \mathbf{C}_k\right) \approx \mathrm{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_k\right) - \mathrm{tr}\left(\sum_{l=1}^{K} \sigma_l^2 \mathbf{C}_W^{-1}\mathbf{C}_l\mathbf{C}_W^{-1}\mathbf{C}_k\right).$$

If we do a similar procedure for the quadratic term on the right-hand side –that is, first obtain the Taylor decomposition of the quadratic form until the first order and apply it to (43), we obtain

$$\begin{array}{c} \mathrm{tr}\left(\left(\mathbf{C}_W + \sum_{l=1}^{K} \sigma_l^2 \mathbf{C}_l\right)^{-1} \mathbf{C}_k \left(\mathbf{C}_W + \sum_{l=1}^{K} \sigma_l^2 \mathbf{C}_l\right)^{-1} \mathbf{R}_X\right) \\ \approx \\ \mathrm{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{R}_X\right) - 2\sum_{l=1}^{K} \sigma_l^2 \mathrm{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{C}_l\mathbf{C}_W^{-1}\mathbf{R}_X\right) \end{array}$$

which leads to expressions of β_{kl} and δ_k in Theorem 9.

L Proof of Theorem 10

The correlation matching approach aims to solve the following minimization problem

$$\hat{\boldsymbol{\sigma}}^2 = \arg\min_{\boldsymbol{\sigma}^2} \left\| \sum_{k=1}^{K} \mathbf{C}_k + \mathbf{C}_W - \mathbf{R}_X \right\|^2 . \tag{91}$$

Consider first the objective function in (91). The Frobenius norm of an Hermitian matrix can be written as $\|\mathbf{R}\|^2 = \mathrm{tr}(\mathbf{R}\mathbf{R})$, and hence

$$\begin{aligned}
\left\| \sum_{k=1}^{K} \mathbf{C}_k + \mathbf{C}_W - \mathbf{R}_X \right\|^2 &= \mathrm{tr}\left(\left(\sum_{k=1}^{K} \mathbf{C}_k + \mathbf{C}_W - \mathbf{R}_X \right) \cdot \left(\sum_{k=1}^{K} \mathbf{C}_k + \mathbf{C}_W - \mathbf{R}_X \right) \right) \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} \sigma_k^2 \sigma_l^2 \mathrm{tr}(\mathbf{C}_k \mathbf{C}_l) + 2 \sum_{k=1}^{K} \mathrm{tr}(\mathbf{C}_k \mathbf{C}_W) - 2 \sum_{k=1}^{K} \mathrm{tr}(\mathbf{C}_k \mathbf{R}_X) \\
&\quad -2\mathrm{tr}(\mathbf{C}_W \mathbf{R}_X) + \mathrm{tr}(\mathbf{C}_W \mathbf{C}_W) + \mathrm{tr}(\mathbf{R}_X \mathbf{R}_X).
\end{aligned}$$

After taking the derivative with respect to σ_k^2, we obtain that

$$2 \sum_{l=1}^{K} \sigma_k^2 \mathrm{tr}(\mathbf{C}_k \mathbf{C}_l) + 2\mathrm{tr}(\mathbf{C}_k \mathbf{C}_W) - 2\mathrm{tr}(\mathbf{C}_k \mathbf{R}_X) = 0,$$

which after some mathematical operations, concludes the proof.

M Assymptotic Performance of GLRT

Consider the case of large sampling depth, i.e., $M \to \infty$. In such a case, we can assume that the sample covariance matrix approaches to the true value, that is, $\mathbf{R}_X \to \mathbf{C}_S + \mathbf{C}_W$. On the one hand, the expressions in Equations (40)–(41) can be equivalently written in a single trace term, by

$$\begin{aligned}
\beta_{kl} &= \mathrm{tr}\left(\mathbf{C}_W^{-1} \mathbf{C}_k \mathbf{C}_W^{-1} \mathbf{C}_l (2\mathbf{C}_W^{-1} \mathbf{R}_X - \mathbf{I}) \right) \\
\delta_k &= \mathrm{tr}\left(\mathbf{C}_W^{-1} \mathbf{C}_k (\mathbf{C}_W^{-1} \mathbf{R}_X - \mathbf{I}) \right).
\end{aligned}$$

The summation in (39) then becomes

$$\begin{aligned}
\sum_{l=1}^{K} \sigma_l^2 \beta_{kl} &= \delta_k \\
\mathrm{tr}\left(\mathbf{C}_W^{-1} \mathbf{C}_k \mathbf{C}_W^{-1} \mathbf{C}_S (2\mathbf{C}_W^{-1} \mathbf{R}_X - \mathbf{I}) \right) &= \delta_k.
\end{aligned}$$

On the other hand, for large data records we have that the terms involving the sample covariance matrix approach to $2\mathbf{C}_W^{-1}\mathbf{R}_X - \mathbf{I} \to 2\mathbf{C}_W^{-1}\mathbf{C}_S + \mathbf{I}$, and also $\mathbf{C}_W^{-1}\mathbf{R}_X - \mathbf{I} \to \mathbf{C}_W^{-1}\mathbf{C}_S$. By plugging these two results into the ML equation, we get the following identity:

$$\mathrm{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\hat{\mathbf{C}}_S(2\mathbf{C}_W^{-1}\mathbf{C}_S+\mathbf{I})\right)=\mathrm{tr}\left(\mathbf{C}_W^{-1}\mathbf{C}_k\mathbf{C}_W^{-1}\mathbf{C}_S\right),$$

where $\hat{\mathbf{C}}_S$ accounts for the ML estimate to be solved. As it can be seen, at low-SNR we have

$$2\mathbf{C}_W^{-1}\mathbf{C}_S=\frac{2}{K}\sum_{l=1}^{K}\frac{\sigma_l^2}{\sigma_W^2}\mathbf{\Sigma}_W^{-1}\mathbf{C}_l\to\mathbf{0}.$$

Therefore, it can be stated that at the low-SNR for large data records, the solution to the ML estimator (39) is consistent with the true value of $\boldsymbol{\sigma}^2$, though uniqueness is not guaranteed.

N Proof of Theorem 11

The Wald test with nuisance parameters is given by [44]

$$\begin{aligned}T_W(\mathbf{X}) &= \left(\hat{\mathbf{\Theta}}_{r1}-\hat{\mathbf{\Theta}}_{r0}\right)^H\\ &\quad\times\left(\left[\boldsymbol{\mathcal{I}}^{-1}(\hat{\mathbf{\Theta}}_1)\right]_{\mathbf{\Theta}_r\mathbf{\Theta}_r}\right)^{-1}\left(\hat{\mathbf{\Theta}}_{r1}-\hat{\mathbf{\Theta}}_{r0}\right),\end{aligned}$$

where in our problem $\hat{\mathbf{\Theta}}_{r1}=\hat{\sigma}_k^2$, $\hat{\mathbf{\Theta}}_{r0}=0$, and $\hat{\mathbf{\Theta}}_1=\hat{\boldsymbol{\sigma}}^2$ is the ML estimate of the unknown parameters, including nuisance parameters, under $\mathcal{H}_{1k}$. After particularizing, we obtain (53).

On the other hand, the Rao test with nuisance parameters is given by [44]

$$\begin{aligned}T_R(\mathbf{X}) &= \left.\frac{\partial\ln p(\mathbf{X}|\mathbf{\Theta})}{\partial\mathbf{\Theta}_r}\right|^H_{\mathbf{\Theta}=\tilde{\mathbf{\Theta}}}\left[\boldsymbol{\mathcal{I}}^{-1}(\tilde{\mathbf{\Theta}})\right]_{\mathbf{\Theta}_r\mathbf{\Theta}_r}\\ &\quad\times\left.\frac{\partial\ln p(\mathbf{X}|\mathbf{\Theta})}{\partial\mathbf{\Theta}_r}\right|_{\mathbf{\Theta}=\tilde{\mathbf{\Theta}}},\end{aligned}$$

where $\tilde{\mathbf{\Theta}}=\hat{\boldsymbol{\sigma}}_{\bar{k}}^2$ is the ML estimate of the unknown parameters under $\mathcal{H}_{0k}$ First, we note that at the low-SNR regime, the outer terms are given by:

$$\begin{aligned}\left.\frac{\partial\ln p(\mathbf{X}|\mathbf{\Theta})}{\partial\mathbf{\Theta}_r}\right|^H_{\mathbf{\Theta}=\tilde{\mathbf{\Theta}}} &= \left.\frac{\partial\ln p(\mathbf{X}|\boldsymbol{\sigma}^2)}{\partial\sigma_k^2}\right|_{\boldsymbol{\sigma}^2=\tilde{\boldsymbol{\sigma}}_{\bar{k}}^2}\\ &= \sum_{l=1}^{K}\sigma_{\bar{k}l}^2\beta_{kl}-\delta_k,\end{aligned}$$

which after being plugged into the general Rao test, concludes the proof.

References

[1] FCC, "Notice of proposed rule making and order," ET Docket No. 03-222, Federal Communications Commission, Dec. 2003.

[2] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. of the IEEE*, vol. 97, pp. 894–914, May 2009.

[3] P. J. Kodolzy, "Cognitive radio fundamentals," in *Proc. Software Defined Radios (SDR)*, Apr. 2005.

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

[5] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Transactions on Communications*, vol. 23, pp. 201–220, Feb. 2005.

[6] "IEEE 802.22 working group on Wireless Regional Area Networks."

[7] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *Proceedings of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 328–337, Nov. 2005.

[8] S. D. Jones, E. Jung, X. Liu, N. Merheb, and I.-J. Wang, "Characterization of spectrum activities in the U. S. public safety band for opportunistic spectrum access," in *Proc. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 137–146, 2007.

[9] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz, "Dynamic frequency hopping communities for efficient IEEE 802.22 operation," *IEEE Communications Magazine*, vol. 45, pp. 80–87, May 2007.

[10] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 2004.

[11] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, pp. 1319–1343, Jun. 2002.

[12] S. Jafar and S. Srinivasa, "Capacity limits of cognitive radio with distributed and dynamic spectral activity," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 529–537, Apr. 2007.

[13] J. G. Proakis, *Digital Communications*. Science/Engineering/Math, McGraw-Hill, 4th ed., August 2000.

[14] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, January 1972.

[15] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, pp. 79–89, May 2007.

[16] M. Timmers, S. Pollin, A. Dejonghe, L. Van der Perre, and F. Catthoor, "A distributed multichannel MAC protocol for multihop cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, pp. 446–459, Jan. 2010.

[17] E. J. Candès and T. Tao, "Decoding with linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[18] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb. 2006.

[19] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb. 2006.

[20] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21–30, Mar. 2008.

[21] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, Jul. 2007.

[22] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 606–617, Dec. 2007.

[23] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, pp. 4655–4666, Dec. 2007.

[24] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0-norm," *IEEE Transactions on Signal Processing*, vol. 57, pp. 289–301, Jan. 2009.

[25] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, Jun. 2002.

[26] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Transactions on Signal Processing*, vol. 57, pp. 843–858, Mar. 2009.

[27] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, pp. 993–1009, Mar. 2009.

[28] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, Jun. 2009. Submitted, http://arxiv.org/abs/arxiv:0906.3173.

[29] M. Mishali and Y. C. Eldar, "From theory to practice: sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, 2010. to appear.

[30] S. Gleichman and Y. C. Eldar, "Blind compressed sensing." http://arxiv1.library.cornell.edu/abs/1002.2586, Feb. 2010.

[31] M. Mishali, Y. C. Eldar, and A. Elron, "Xampling – part I: Practice." http://arxiv.org/abs/0911.0519, Nov. 2009.

[32] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: analog to digital at sub-Nyquist rates." http://arxiv.org/abs/0912.2495, Dec. 2009.

[33] G. Reeves and M. Gastpar, "'Compressed' compressed sensing." arXiv:1001.4295v1, Jan. 2010.

[34] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill Companies, Aug. 1979.

[35] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[36] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, pp. 32–39, Apr. 2008.

[37] H. V. Poor, *An introduction to signal detection and estimation*. New York, NY: Springer-Verlag, 1994.

[38] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *Proc. IEEE International Symposium of New Frontiers on Dynamic Spectrum Access (DySPAN)*, pp. 151–159, Nov. 2005.

[39] P. Sutton, K. Nolan, and L. Doyle, "Cyclostationary signatures in practical cognitive radio applications," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 13–24, Jan. 2008.

[40] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, pp. 849–877, May 2009.

[41] M. Hernandez, M. Rojas, and P. Stoica, "New spectral estimation based on filterbank for spectrum sensing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3509–3512, Apr. 2008.

[42] A. I. Pérez-Neira, M. A. Lagunas, M. A. Rojas, and P. Stoica, "Correlation matching approach for spectrum sensing in open spectrum communications," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4823–4836, Dec. 2009.

[43] W. A. Gardner, *Statistical Spectral Analysis. A Non Probabilistic Theory*. Prentice-Hall, 1988.

[44] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 2 (Detection Theory). Upper Saddle River, NJ: Prentice Hall, 1998.

[45] J. Ma and Y. Li, "A probability-based spectrum sensing scheme for cognitive radio," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 3416–3420, May 2008.

[46] T. J. Lim, R. Zhang, Y. C. Liang, and Y. Zeng, "GLRT-based spectrum sensing for cognitive radio," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, pp. 4391–4395, Dec. 2008.

[47] J. Font-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information," *IEEE Transactions on Communications*, 2010. Accepted.

[48] C.-H. Hwang, G.-L. Lai, and S.-C. Chen, "Spectrum sensing in wideband OFDM cognitive radios," *IEEE Transactions on Signal Processing*, vol. 58, pp. 709–719, Feb. 2010.

[49] Z. Tian, E. Blasch, W. Li, G. Chen, and X. Li, "Performance evaluation of distributed compressed wideband sensing for cognitive radio networks," in *Proc. International Conference on Information Fusion*, pp. 1–8, Jul. 2008.

[50] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, pp. 1–5, Dec. 2008.

[51] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, pp. 1357–1360, Apr. 2007.

[52] X. Chen, L. Zhao, and J. Li, "A modified spectrum sensing method for wideband cognitive radio based on compressive sensing," in *Proc. International Conference on Communications and Networking (ChinaCOM)*, pp. 1–5, Aug. 2009.

[53] Y. Wang, A. Pandharipande, Y. L. Polo, and G. Leus, "Distributed compressive wide-band spectrum sensing," in *Proc. Information Theory and Applications Workshop (ITA)*, pp. 178–183, Feb. 2009.

[54] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2337–2340, Apr. 2009.

[55] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1847–1862, Mar. 2010.

[56] S. Geirhofer, L. Tong, and B. M. Sadler, "Measurement-based models enabling cognitive medium access in wlan bands," Tech. Rep. ACSP TR-02-07-02, Cornell University, Feb. 2007.

[57] S. Benedetto and E. Biglieri, *Principles of Digital Communications*. Information Technology: Transmission, Processing, and Storage, New York: Kluwer Academic / Plenum Publishers, 1999.

[58] S. Weber, X. Yang, J. Andrews, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Transactions on Information Theory*, vol. 51, pp. 4091–4102, Dec. 2005.

[59] S. Weber, J. G. Andrews, and N. Jindal, "A tutorial on transmission capacity." http://arxiv.org/abs/0809.0016, Aug. 2009.

[60] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels. II. Outage capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1103–1127, Mar. 2001.

[61] L. Li and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1326–1347, Apr. 2005.

[62] M. B. Pursley, "The role of spread spectrum in packet radio networks," *Proc. of the IEEE*, vol. 75, pp. 116–134, Jan. 1987.

[63] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communication aspects,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2019–2692, Oct. 1999.

[64] G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels,” *IEEE Transactions on Information Theory*, vol. 45, pp. 1468–1489, Jul. 1999.

[65] L. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[66] G. Kaplan and S. Shamai, “Error probabilities for the block-fading gaussian channel,” *Arch. Elek. Ubertrangung*, vol. 49, no. 4, pp. 192–205, 1995.

[67] W. Choi and J. Andrews, “Spatial multiplexing in cellular MIMO-CDMA systems with linear receivers: Outage probability and capacity,” *IEEE Transactions on Wireless Communications*, vol. 6, pp. 2612–2621, Jul. 2007.

[68] L. Kleinrock and J. A. Silvester, “Optimum transmission radii in packet radio networks or why six is a magic number,” in *Proc. National Telecommunications Conference*, Dec. 1978.

[69] S. Musa and W. Wasylkiwskyj, “Co-channel interference of spread spectrum systems in a multiple user environment,” *IEEE Transactions on Communications*, vol. 26, pp. 1405–1413, Oct. 1978.

[70] F. Bacceli and B. Blaszczyszyn, *Stochastic geometry and wireless networks*, vol. 3-4 of *Foundations and Trends in Networking*. Now Publishers, Inc., 2009.

[71] S. Weber, X. Yang, G. de Veciana, and J. G. Andrews, “Transmission capacity of CDMA ad hoc networks,” in *Proc. International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, (Australia), pp. 245–249, 2004.

[72] S. Weber, J. Andrews, X. Yang, and G. de Veciana, “Transmission capacity of wireless ad hoc networks with successive interference cancellation,” *IEEE Transactions on Information Theory*, vol. 53, pp. 2799–2814, Aug. 2007.

[73] N. Jindal, J. G. Andrews, and S. Weber, “Bandwidth partitioning in decentralized wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 5408–5419, Dec. 2008.

[74] V. Chandrasekhar and J. Andrews, “Spectrum allocation in tiered cellular networks,” *IEEE Transactions on Communications*, vol. 57, pp. 3059 –3068, Oct. 2009.

[75] W. Weber, J. Andrews, and N. Jindal, “The effect of fading, channel inversion, and threshold scheduling on ad hoc networks,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4127–4149, Nov. 2007.

[76] N. Jindal, S. Weber, and J. G. Andrews, “Fractional power control for decentralized wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 5482–5492, Dec. 200.

[77] A. Hunter, J. Andrews, and S. Weber, “The transmission capacity of ad hoc networks with spatial diversity,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 5058–5071, Dec. 2008.

[78] J. G. Andrews, S. Weber, and M. Haenggi, “Ad hoc networks: to spread or not to spread,” *IEEE Communications Magazine*, vol. 45, no. 12, pp. 84–91, 2007.

[79] R. Menon, R. M. Buehrer, and J. H. Reed, “Outage probability based comparison of underlay and overlay spectrum sharing techniques,” in *Proc. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 10 –109, Nov. 2005.

[80] S. Geirhofer, L. Tong, and B. M. Sadler, “Measurement-based models for cognitive medium access in WLAN bands,” Technical Report ACSP-TR-02-07-02, Cornell University, Adaptive Communications and Signal Processing Group (ACSP), Feb. 2007. http://acsp.ece.cornell.edu/papers/ACSP-TR-02-07-02.pdf.

[81] S. Geirhofer, J. Z. Sun, L. Tong, and B. M. Sadler, “Cognitive frequency hopping based on interference prediction: theory and experimental results,” *Mobile Computing and Communications Review*, vol. 13, pp. 49–61, Mar. 2009.

[82] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 589–600, Apr. 2007.

[83] S. Geirhofer, L. Tong, and B. Sadler, “Cognitive medium access: Constraining interference based on experimental models,” *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 95–105, Jan. 2008.

[84] Q. Zhao, S. Geirhofer, L. Tong, and B. Sadler, “Opportunistic spectrum access via periodic channel sensing,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 785–796, Feb. 2008.

[85] E. Altman, *Constrained Markov Decision Process*. Chapman & Hall/CRC, 1999.

[86] K. Nishimori, R. D. Taranto, H. Yomo, P. Popovski, Y. Takatori, R. Prasad, and S. Kubota, “Spacial opportunity for cognitive radio systems with heterogeneous path loss conditions,” in *Proc. IEEE Vehicular Technology Conference*, pp. 2631–2635, Aug. 2007.

[87] J. Riihijärvi and P. Mähönen, “Exploiting spatial statistics of primary and secondary users towards improved cognitive radio networks,” in *Proc. International Conference on Cognitive Radio Oriented Wireless Network Communications*, pp. 1–7, May 2008.

[88] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1 (Estimation Theory). Upper Saddle River, NJ: Prentice Hall, 1998.

[89] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. of the IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[90] A. de Moivre, *The Doctrine of Chances*. London: H. Woodfall, 2nd ed., 1738.

[91] G. Moustakides, “Finite sample size optimality of GLR tests,” *IEEE Transactions on Information Theory*, 2009. Submitted, http://arxiv.org/abs/0903.3795.

[92] B. Porat, *Digital Processing of Random Signals. Theory and Methods*. Dover Publications, Feb. 2008.

[93] A. Goldsmith and M. Medard, "Capacity of time-varying channels with causal channel side information," *IEEE Transactions on Information Theory*, vol. 53, pp. 881–899, Mar. 2007.

[94] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 18, pp. 725–730, Nov. 1976.

[95] A. Goldsmith, *Wireless communications*. Cambridge University Press, Aug, 2005.

[96] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, (Budapest, Hungary), pp. 23–52, Sep. 1971.

[97] H. Liao, *Multiple Access Channels*. PhD thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[98] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, Jul. 2003.

[99] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 1875–1892, Sep. 2004.

[100] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, Oct. 2003.

[101] C.-H. F. Fung, W. Yu, and T. J. Lim, "Precoding for the multi-antenna downlink: multi-user SNR gap and optimal user ordering," *IEEE Transactions on Communications*, vol. 55, pp. 188–197, Jan. 2007.

[102] T. Sartenaer, L. Vandendorpe, and J. Louveaux, "Balanced capacity of wire line multi-user channels," *IEEE Transactions on Communications*, vol. 53, pp. 2029–2042, Dec. 2005.

[103] A. Kalakech, J. Louveaux, and L. Vandendorpe, "Applying the balanced capacity concept to DSL systems," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 2717–2721, Jun. 2007.

[104] Y. Sun, "Spectral efficiency and optimal medium access control of random access systems over large random spreading CDMA," *IEEE Transactions on Communications*, vol. 57, pp. 1256–1260, May 2009.

[105] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, pp. 622–640, Mar. 2009.

[106] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, J. Weaver, L. A., and C. E. Wheatley, "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 303–312, May 1991.

[107] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proc. Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), pp. 481–492, University of California Press, 1950.