Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

PHILIPS
sense and simplicity

M.Sc. Thesis

Compressive Wideband Spectrum Sensing for Cognitive Radio Applications

Yvan Lamelas Polo

Abstract

It has been widely recognized that utilization of radio spectrum by licensed wireless systems, e.g., TV broadcasting, aeronautical telemetry, is quite low. In particular, at any given time and spatial region, there are frequency bands where there is no signal occupancy. There has been recent interest in improving spectrum utilization by permitting secondary usage using cognitive radios. Cognitive radios use spectrum sensing to determine frequency bands that are vacant of licensed signal transmissions and transmit on such portions to meet regulatory constraints of avoiding harmful interference to licensed systems. Future cognitive radios will be capable of scanning a wide band of frequencies, in the order of a few GHz, and employ adaptive waveforms for transmission depending on the estimated spectrum of licensed systems. In this thesis, we address the problem of estimating the spectrum of the wide-band signal received at the cognitive radio sensing receiver using compressive sampling coupled with a multiband spectrum detector to determine the spectrum occupancy of the licensed system. Since individual cognitive radios might not be able to reliably detect weak primary signals due to channel fading/shadowing, we also propose a distributed compressive scheme based on joint recovery of the license occupancy for application scenarios involving geographically distributed radios. In such a distributed approach, the spectrum occupancy is determined by the joint work of cognitive radios (exploiting spatial diversity), as opposed to being determined individually by each cognitive radio.

TU Delft

Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Compressive Wideband Spectrum Sensing for Cognitive Radio Applications

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Yvan Lamelas Polo
born in Nyon, Switzerland

This work was performed in:

Distributed Sensor Systems Group
Philips Research Europe
High Tech Campus
5656 AE Eindhoven, The Netherlands

Delft University of Technology
Department of
Microelectronics & Computer Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Compressive Wideband Spectrum Sensing for Cognitive Radio Applications"** by **Yvan Lamelas Polo** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: November 28, 2008

Chairman: ______________________________
Prof. dr. ir. A.-J. van der Veen, CAS, TU Delft

Advisors: ______________________________
Dr. ir. Ashish Pandharipande, Philips Research Europe

____
Dr. ir. Geert Leus, CAS, TU Delft

Committee Members: ______________________________
Prof. dr. ir. Hakan Deliç, Boğaziçi University, Istanbul

Abstract

It has been widely recognized that utilization of radio spectrum by licensed wireless systems, e.g., TV broadcasting, aeronautical telemetry, is quite low. In particular, at any given time and spatial region, there are frequency bands where there is no signal occupancy. There has been recent interest in improving spectrum utilization by permitting secondary usage using cognitive radios. Cognitive radios use spectrum sensing to determine frequency bands that are vacant of licensed signal transmissions and transmit on such portions to meet regulatory constraints of avoiding harmful interference to licensed systems. Future cognitive radios will be capable of scanning a wide band of frequencies, in the order of a few GHz, and employ adaptive waveforms for transmission depending on the estimated spectrum of licensed systems. In this thesis, we address the problem of estimating the spectrum of the wide-band signal received at the cognitive radio sensing receiver using compressive sampling coupled with a multiband spectrum detector to determine the spectrum occupancy of the licensed system. Since individual cognitive radios might not be able to reliably detect weak primary signals due to channel fading/shadowing, we also propose a distributed compressive scheme based on joint recovery of the license occupancy for application scenarios involving geographically distributed radios. In such a distributed approach, the spectrum occupancy is determined by the joint work of cognitive radios (exploiting spatial diversity), as opposed to being determined individually by each cognitive radio.

Acknowledgments

Although my name appears as the sole author of the thesis, in reality a number of people are responsible for the work. First and foremost, I would like to thank my thesis advisors, Ashish Pandharipande, and Geert Leus. They introduced me to the topics of compressive sensing and cognitive radio and they steered my research efforts to attack interesting and potentially solvable problems. I must thank Ashish, for his titanic efforts in making my papers and my thesis readable, and teaching me the process of good scientific writing. Many thanks Geert, for your always positive attitude and for helping me to mend my notation inconsistencies. I cannot forget to thank Ying Wang, the great deal of help from her ranges from explanations of obscure mathematical ideas, to tips for effective programming. Many thanks Ying, for reading my essays, and attending most of the progression meetings with Ashish giving always valuable criticism. Working with Ashish, Geert and Ying has been very inspiring: their energy, dedication, creative ideas and excellent organization giving me invaluable guidance. I would also like to thank Zhi Tian and Frans Willems for the discussions on my work. I acknowledge Professor Hakan Deliç and A.-J. van der Veen for their participation in my MSc thesis committee. The outcomes from the last year of interaction with these people did not get reflected in the thesis, but I feel they deserve to be mentioned anyway. I could not have wished for better collaborators and coaches. Your contributions, detailed comments and insights have been of great value for me. I am also grateful for the opportunity to be part of a research group at Philips.

My warmest thanks go to my friends, very specially, Miguel Alvarez, Iria Varela, Blanca Barrientos, Manel Herrer and Daniela Deiana for making me forget for brief periods of time about my work and my thesis.

Finally, and most importantly, I would like to express gratitude to my family. Even thousands of kilometers apart, they have been present through every step of my life, providing support in difficult times. They have been a constant source of inspiration, and this thesis is dedicated to them.

Yvan Lamelas Polo
Eindhoven
November 28, 2008

Contents

List of Figures

Introduction

1

In this thesis we consider the problem of spectrum sensing for cognitive radio applications, and present a new approach based on compressive sampling. The purpose of this chapter is to introduce the problem addressed in the thesis, motivate the need for a new approach, and describe our main contributions and the organization of the thesis.

1.1 Motivation: spectrum sensing for spectrum sharing

The development of wireless technologies has rapidly increased the demand for spectrum resources. However, most of the spectrum has already been allocated to licensed users or primary users (PUs), especially in the frequency below a few GHz. The National Telecommunications and Information Administration's (NTIA) frequency allocation chart in the United States (Fig. 1.1) indicates overlapping allocations over all of the frequency bands, which reinforces the scarcity mindset. Under this static frequency allocation wireless systems are regulated through fixed spectrum assignments, operating frequencies and bandwidths, with constraints on power emission that limits their range. Due to these constraints, most communications systems are designed so that they achieve the best possible spectrum efficiency within the assigned bandwidth using sophisticated modulation, coding, multiple antennas and other techniques.

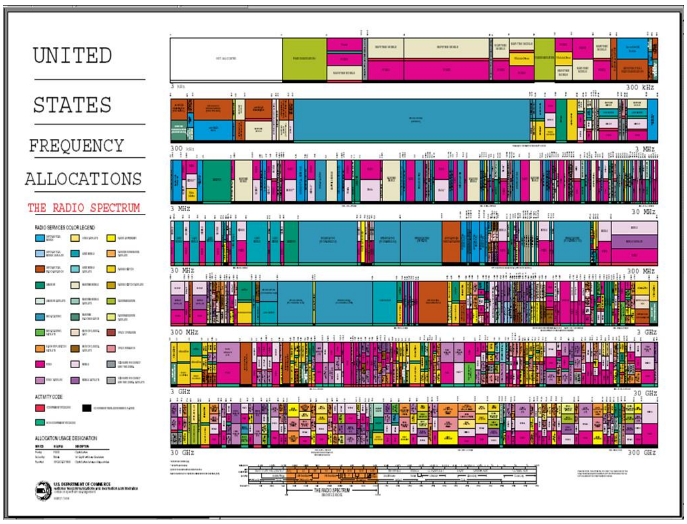

Figure 1.1: The NTIA's frequency allocation chart.

While the current spectrum allocation leaves no available bandwidth for future wireless systems, actual measurements of spectrum utilization show that many assigned bands are not being used at every location and time [2]. The underutilization of the electromagnetic spectrum leads us to think in terms of spectrum holes, i.e., bands of frequencies assigned to a primary user, but at a particular time and specific geographic location, the band is not being utilized by that user. Figures 1.2 and 1.3 show some measurement results showing a typical utilization of spectrum resources. We can see that a greater percentage of the spectrum is available at higher frequencies. This result makes sense intuitively as, for a given transmitter power level, a wireless signal will propagate further at a lower frequency, leading to a higher observed utilization at any particular location. Studies carried out by the FCC's (Federal Communications Commision) Spectrum Policy Task Force reported the vast temporal and geographic variations in the usage of allocated spectrum with utilization ranging from 15% to 85% [3]. A recent study conducted by Shared Spectrum [1] shows that the average spectrum occupancy in the frequency band from 30 MHz to 3000 MHz over multiple locations is merely 5.2%. The maximum occupancy is about 13% in New York city. These measurements seriously question the suitability of the current regulatory regime and possibly provide the opportunity to solve the spectrum bottleneck.

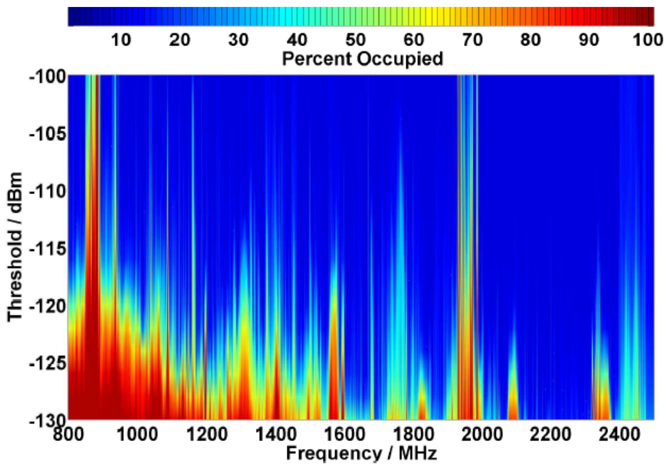

Figure 1.2: Measured spectrum utilization vs. frequency for the measurements recorded in Annapolis [1].

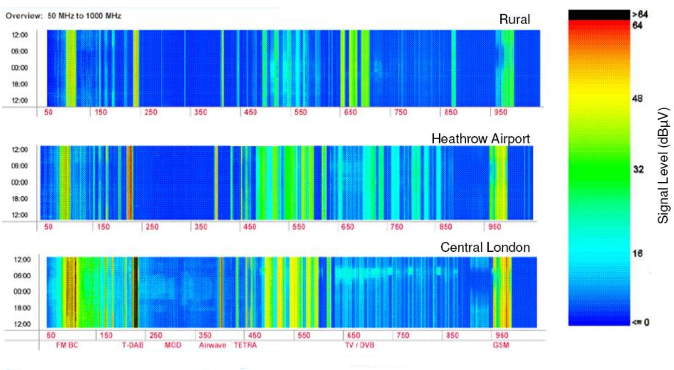

Figure 1.3: Measurement of spectrum utilization.

Unfortunately, creating a new spectrum allocation chart based on the usage distribution is not only impractical but also inefficient, because it is not possible to predict and optimize an allocation that would suit all current and future wireless systems. Furthermore, any change in the spectrum allocation could create an opposition from the current users/owners of the spectrum. Therefore, the solution to this problem should preserve rights and access priorities of "primary users".

In order to solve the conflicts between spectrum scarcity and spectrum underutilization, cognitive radio (CR) technology was recently proposed [4]. In IEEE 802.22, the CR technique is introduced for the standardization of wireless regional area networks (WRAN) to use frequency resources, which were originally allocated for broadcasting ($54 \sim 862MHz$). In [20] a CR is defined as an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environmental and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit power, carrier frequency, and modulation strategy) in real-time, with two primary objectives in mind:

- highly reliable communications whenever and wherever needed;
- efficient utilization of the radio spectrum.

Six key words stand out from this definition: awareness, intelligence, learning, adaptivity, reliability, and efficiency. Consequently, the CR will play the leading role in the transition from fixed allocation assignments to dynamic spectrum allocation.

In order to protect the PUs from unlicensed users or secondary users (SUs), spectrum sensing is a key function to decide whether a frequency band is empty or not. As explained before, a cognitive radio is designed to be aware of and sensitive to the changes in its surroundings. Therefore, the SUs should monitor licensed bands, and opportunistically transmit whenever no primary signal is detected. Consequently spectrum sensing may be identified as a key enabling functionality to ensure that a CR would not interfere with primary users. In the rest of this thesis we address some of the major challenges behind wideband spectrum sensing.

1.2 Spectrum sensing challenges

In short, CRs as secondary users of license bands have to dynamically sense the radio-spectrum environment and rapidly tune their transmitter parameters to efficiently utilize the available spectrum. The critical problem is the need to process multi-gigahertz wide bandwidth (challenging traditional spectral estimation methods operating at or above Nyquist rate) and reliably detect presence of primary users. We next analyze this problem by splitting it up into two challenging tasks to be achieved:

- *Detection capability* [7]: In general, PUs have not been very receptive at the idea of CRs and opportunistic spectrum sharing. In particular, they are concerned that CRs will harmfully interfere with their operation. However, in this argument, it is not very well understood what is considered harmful interference. A first example of harmful interference is when a CR may not be able to reliably detect a PU signal and therefore may start sending although the PU is using that frequency band. This is the classic “hidden terminal problem” in wireless networks where a receiver is unable to “hear” the transmitter and starts its own transmission, thereby interfering with the intended receiver of the transmission. A second example is when a cognitive radio is using a frequency band that was deemed free by the sensing process but may not be able to reliably detect that a PU has reappeared. Therefore, it may not vacate the frequency band quickly enough and therefore continue to send creating harmful interference to the primary user’s transmission. From these two examples, we can see that there are clearly two requirements on a CR sensing receiver that influence the amount of harmful interference. In the first example, it is evident that receiver sensitivity plays a key role in reliable detection. In addition, the objective of resource allocation is to meet interference constraints of primary systems. These interference constraints are met through adaptive transmit power allocation based on spectrum sensing measurements. In principle, if a radio can meet specified sensitivity levels, it should be allowed to transmit higher power levels as it is located far away from the protected radius. Based on this rationale, there is an explicit trade-off between the sensing sensitivity and allowable transmission power. In the second example, the sensing time needed to meet the required sensitivity is another requirement for sensing performance. The sensing interval requirement presents the maximum time a cognitive radio sensor could spend for primary user detection.

- *Implementation of wideband front-end and sampling circuitry*: The wideband radio-frequency (RF) signal presented at the antenna of a wideband front-end includes signals from close and far transmitters. One of the main limitations in a radio front-end’s ability to detect small signals is its dynamic range, which also dictates the requirement for the number of bits in the analog-to-digital converter (ADC). The wideband sensing requires a multi-GHz speed ADC together with high resolution. In [7], the dynamic range problem is addressed by proposing to filter the signals in the spatial domain with multiple antennas. In this thesis we investigate new signal processing techniques that can relax the challenging requirements for the ADC. In particular, we make use of an emerging theory named compressive sampling (CS). CS is a method for acquisition of sparse signals at rates significantly lower than the Nyquist rate; signal reconstruction is a solution to an optimization problem.

1.3 Outline and contributions

Before describing the content of the thesis chapter by chapter, we briefly summarize our main contributions. The first major contribution is the development of an acquisition scheme for compressive sensing at local CR sensing receivers. In this framework we estimate the power spectrum density (PSD) of a wideband signal from compressive sub-Nyquist rate measurements, as the solution of an edge detector optimization problem and we then perform energy based detection on the compressive estimate of the

PSD to detect channel occupancy. However, if one CR does not see energy in a particular band, it cannot assume that the channel is idle (i.e., hidden terminal problem). Hence, the second contribution of the thesis is the extension to a distributed cognitive wireless network scenario. We propose two distributed architectures extending the PSD estimation based on compressive edge detection beyond a single cognitive radio sensing receiver. The third contribution is the development of a joint recovery algorithm for multiple measurement vectors under common sparsity constraints.

Chapter 2: Background discussion

In this chapter we present a literature review of some specific detection and estimation techniques. We end the chapter by describing some of the limitations of these techniques thereby motivating the need for our compressive estimation method.

Chapter 3: Compressive Wideband Sensing

This chapter contains the first contribution of the thesis. We start by giving a brief overview of a compressive sensing technique existing in current literature. We discuss its limitations and we then propose an acquisition scheme coupled with a compressive edge detector for PSD estimation. We evaluate the performance of the proposed estimator in terms of the mean square error (MSE). Finally, a PSD based energy detector is proposed to find channel occupancy. Performance of the detector is evaluated in terms of probability of detection and probability of false alarm.

Chapter 4: Distributed Spectrum Sensing

In this chapter we address the lack of reliability when detection is only performed at a local sensing receiver. Consequently, we propose two distributed architectures to provide spatial diversity gain under unfavorable channel conditions. Algorithms based on independent CS recovery and joint CS recovery are provided and their detection performance compared.

Chapter 5: Practical Performance Issues

This chapter is devoted to the analysis of the practical implementation issues of the techniques developed in the previous chapters. First we describe the effects of the different parameters involved in spectrum sensing. Next, we discuss how the assumptions made in the thesis affect implementation. We then provide an overview of the actual techniques proposed in the literature to implement analog to information (A2I) converters and we finally examine a detection algorithm where the intermediate stage of estimating the PSD is avoided.

Chapter 6: Conclusions and Further work

This chapter summarizes the main ideas of the thesis and gives suggestions for further research in the area.

Background discussion

2

Since CRs are secondary users of unoccupied spectrum they do not have *a priori* right to any frequency band. Their communication is strictly conditional on the reliable detection of PU transmissions in their vicinity. As a result, CRs must operate in a much wider frequency bandwidth than conventional radios which spans multiple PU bands and CRs must perform frequent measurements of PUs' activity through spectrum sensing. The most autonomous and flexible approach which could be used to check the presence of PU signals, is based on measurements of the actual occupancy at a given location and time. In this chapter we describe the basic detection/estimation groundwork for wideband sensing. The literature of wideband sensing for CR networks is very limited. An early approach is to use a tunable narrowband bandpass filer (BPF) at the RF frond-end to sense one narrow frequency band at a time [9], over which the existing narrowband spectrum sensing techniques discussed in Section 2.1 can be applied. In order to operate over multiple frequency bands at a time, the RF front-end requires a wideband architecture followed by a high-speed DSP and the spectrum sensing usually involves the estimation of the PSD of the wideband signal. Classical estimation techniques of the PSD are discussed in Section 2.2. Alternatively, multiple narrowband BPFs [10] may be employed, but this architecture requires an increased number of components and the filter range of each BPF is preset. In the wideband regime, a major challenge stems from the high RF signal acquisition costs of current ADC hardware technology. We shall close this chapter with a new emergent promising technology called compressive sampling which allows for sub-Nyquist rate sampling of sparse signals alleviating the sampling burden and energy consumption.

2.1 Detection techniques

In detection theory for CR systems we want to determine whether a PU is present or not. This results in a binary hypothesis test, where two cases (hypotheses) are stated and the algorithm has to decide which one is (most likely) true. The classical detection problem is to distinguish between the hypotheses

$$\begin{aligned}
&\mathcal{H}_0 : x[n] = w[n] && n = 0, 1, \ldots, N-1 \\
&\mathcal{H}_1 : x[n] = s[n] + w[n] && n = 0, 1, \ldots, N-1
\end{aligned} \tag{2.1}$$

where $x[n]$ is the signal received by the cognitive user, $s[n]$ is the noiseless received signal when the PU is present and $w[n]$ is the additive white gaussian noise (AWGN).

The performance of a detector can be characterized by its probability of correct detection p_d and false alarm rate p_{fa}. When working with hypothesis testing, there are a number of different errors that can occur

1. Type I: decide $\mathcal{H}_1$ but $\mathcal{H}_0$ is true (*false alarm*).

2. Type II: decide $\mathcal{H}_0$ but $\mathcal{H}_1$ is true (*missed detection*).

Four schemes are generally used for detection according to the hypothesis model:

1. When the information of the PU signal is known to the cognitive user, the optimal detector in stationary Gaussian noise is the *matched filter* (coherent detection) since it maximizes the received signal-to-noise ratio (SNR). While the main advantage of the matched filter is that it requires less time to achieve high processing gain due to coherency, implementing this type of coherent detector is difficult since a SU would need extra dedicated circuitry to achieve carrier synchronization with each type of license user. Moreover there may be cases in practice where matched filtering is ruled out due to the lack of knowledge about PUs. In [11] and [12] compressive detectors using matched filtering have been proposed.

2. If the receiver cannot gather any information about the PU signal, the optimum detector is an *energy detector* (non-coherent detection). Since it is easy to implement, and also it is the most general technique since it applies to any signal type, recent work on detection of the PU has generally adopted the energy detector [13] [14]. It requires minimum information about the signal, including only signal bandwidth and carrier frequency. However, the performance of the energy detector is susceptible to an uncertainty in the noise power. Another shortcoming is that the energy detector cannot differentiate signal types but can only determine the presence of the signal. Thus, the energy detector is prone to false detection triggered by the unintended signals.

3. While energy detection is a fairly general approach, it neglects the presence of deterministic signals like pilots that PUs embed in their transmissions in order to perform synchronization and acquisition. Hence *pilot detection* represents a different technique which relies on the fact that the power of the known pilot tone is typically 1% to 10% of the total transmitted power. One special case of a pilot signal, frequently present in PU broadcast systems, is a sinewave tone used for receiver synchronization. This method has been suggested for CR sensing receivers in [15] and [16].

4. An alternative detection method is the *cyclostationary feature detection* [17]. Modulated signals are in general coupled with sine wave carriers, pulse trains, repeating spreading, hopping sequences, or cyclic prefixes, which result in built-in-periodicity. These modulated signals are characterized as cyclostationary signals since their mean and autocorrelation exhibit periodicity. These features are detected by analyzing a spectral correlation function. The main advantage of the spectral correlation function is that it differentiates the noise energy from modulated signal energy, which is a result of the fact that the noise is a wide-sense stationary signal with no correlation, while modulated signals are cyclostationary with spectral correlation due to the embedded redundancy of signal periodicity. Therefore, a cyclostationary feature detector can perform better than the energy detector in discriminating against noise due to its robustness to the uncertainty in

noise power. However, it is computationally complex and requires prior knowledge about the PU signal structure and significantly long observation time.

2.2 Estimation techniques

The goal of *spectral estimation* is to describe the distribution over frequency of the power contained in a signal, based on a finite set of data. Estimation of power spectra is useful in a variety of applications, including the detection of signals buried in wideband noise.

The PSD of a stationary random process $x(n)$ is mathematically related to the correlation sequence $r_x(n)$ by the discrete time Fourier transform. This is given by

$$S_x(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} r_x(n)e^{-j\omega n} \qquad -\pi < \omega \leq \pi \tag{2.2}$$

with

$$r_x(n) = E[x^*(m)x(m+n)]. \tag{2.3}$$

The average power of a signal over a particular frequency band $[\omega_1, \omega_2]$, $0 \leq \omega_1 < \omega_2 \leq \pi$, can be found by integrating the PSD over that band

$$P_{[\omega_1,\omega_2]} = \int_{\omega_1}^{\omega_2} S_x(e^{j\omega})d\omega. \tag{2.4}$$

We can see from the above expression that $S_x(e^{j\omega})$ represents the power content of a signal in an infinitesimal frequency band, which is why it is called the power spectral density.

The main methods for wideband spectrum estimation can be divided into non-parametric and parametric methods

1. *Nonparametric methods* are those in which the PSD is estimated directly from the signal itself. The simplest of such methods is the periodogram. An improved version of the periodogram is Welch's method. A more modern nonparametric technique is the multitaper method.

2. *Parametric methods* are those in which the PSD is estimated from a signal that is assumed to be the output of a linear system driven by white noise. Examples are the Yule-Walker autoregressive (AR) method and the Burg method. These methods estimate the PSD by first estimating the parameters (coefficients) of the linear system that hypothetically generates the signal. They tend to produce better results than classical nonparametric methods when the data length of the available signal is relatively short.

Following [18] Appendix A presents a short review of parametric and nonparametric methods showing simulations of a practical signal from a cognitive radio sensing receiver perspective.

2.3 Compressive sampling

Spectrum sensing in the wideband regime faces considerable technical challenges. A major implementation challenge lies in the very high sampling rates required by conventional spectral estimation methods which have to operate at or above the Nyquist rate.

The Nyquist sampling theorem tells us that in order not to lose information when uniformly sampling a signal we must sample at least two times faster than its bandwidth. In many applications, the Nyquist rate can be so high that we end up with too many samples and must compress in order to store or transmit them. In other applications, increasing the sampling rate or density beyond the current state-of-the-art is very expensive.

In this section, an emerging field called *compressive sampling* (CS) will be explained. CS builds on the works of Candes, Romberg, and Tao [23] [24] and Donoho [22], who showed that if a signal has a sparse representation in one basis then it can be recovered from a small number of projections onto a second basis that is incoherent with the first.

Nyquist-rate sampling completely describes a signal by exploiting its bandlimitedness. The goal of CS is to reduce the number of measurements required to completely describe a signal by exploiting its compressibility. The difference will be that the measurements are not point samples any more but more general linear functionals of the signal.

Let us consider a discrete-time signal $\mathbf{x}$, which we view as an $N \times 1$ column vector with elements $x[n]$, $n = 1, 2, , N$. Any signal can be represented in terms of a basis of $N \times 1$ vectors $\{\boldsymbol{\psi}_i\}_{i=1}^{N}$. Forming the $N \times N$ basis matrix

$$\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2 \mid \ \ldots \ \mid \boldsymbol{\psi}_N] \tag{2.5}$$

by stacking the vectors $\{\boldsymbol{\psi}_i\}$ as columns, we can express any signal $\mathbf{x}$ as

$$\mathbf{x} = \sum_{i=1}^{N} s_i \boldsymbol{\psi}_i \tag{2.6}$$

or

$$\mathbf{x} = \boldsymbol{\Psi}\mathbf{s} \tag{2.7}$$

where $\mathbf{s}$ is the $N \times 1$ column vector of weighting coefficients. Clearly, $\mathbf{x}$ and $\mathbf{s}$ are equivalent representations of the same signal, with $\mathbf{x}$ in the time domain and $\mathbf{s}$ in the $\boldsymbol{\Psi}$ domain.

We focus on signals that have a sparse representation, where $\mathbf{x}$ is a linear combination of just K basis vectors, with $K \ll N$. That is, only K of the s_i coefficients in (2.6) are nonzero. Sparsity is motivated by the fact that many natural and manmade signals are compressible in the sense that there exists a basis $\boldsymbol{\Psi}$ where the representation has just a few large coefficients and many small coefficients and represents a requirement in order to apply the CS framework.

Let's consider the linear measurement process that computes $M < N$ inner products between $\mathbf{x}$ and a collection of vectors $\{\boldsymbol{\phi}_j\}_{j=1}^{M}$ as in

$$y_j = \langle \mathbf{x}, \boldsymbol{\phi}_j \rangle. \tag{2.8}$$

Stacking the measurements y_j into the $M \times 1$ vector $\mathbf{y}$ and the measurement vectors $\boldsymbol{\phi}_j^T$ as rows into an $M \times N$ matrix $\mathbf{\Phi}$ and substituting this in (2.6), we can write

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi s} = \mathbf{\Theta s} \tag{2.9}$$

where $\mathbf{\Theta}$ is an $M \times N$ matrix. It should be noted that the measurement process is non-adaptive; that is, $\mathbf{\Phi}$ does not depend in any way on the signal $\mathbf{x}$. Figure 2.1 shows an illustration of (2.9).

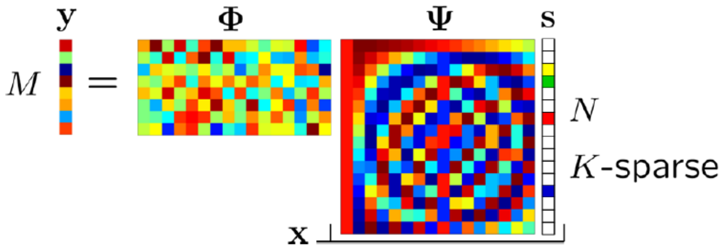

Figure 2.1: Compressive sensing measurement process [60].

The solution consists of two steps. In the first step, a stable measurement matrix $\mathbf{\Phi}$ is designed which ensures that the information is not damaged by the dimensionality reduction from $\mathbf{x} \in \mathcal{C}^N$ down to $\mathbf{y} \in \mathcal{C}^M$. In the second step, a reconstruction algorithm is developed to recover $\mathbf{x}$ from the measurements $\mathbf{y}$.

2.3.1 Stable measurement matrix

The goal is to make M measurements from which the length-N signal $\mathbf{x}$ could be reconstructed, or equivalently its sparse coefficient vector $\mathbf{s}$ in the basis $\mathbf{\Psi}$. Clearly reconstruction will not be possible if the measurement process damages the information in $\mathbf{x}$. Hence, since the measurement process is linear and defined in terms of the matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, solving for $\mathbf{s}$ given $\mathbf{y}$ in (2.9) is a linear algebra problem, with $M < N$, i.e., fewer equations than unknowns, resulting in an infinite number of solutions (ill-posed).

However the K-sparsity of $\mathbf{s}$ comes to the rescue and an intuitive approach to ensure the solution is that the measurement matrix $\mathbf{\Phi}$ is incoherent with the sparsifying basis $\mathbf{\Psi}$ [23] in the sense that the vectors $\{\boldsymbol{\phi}_j\}_{j=1}^M$ cannot sparsely represent the vectors $\{\boldsymbol{\psi}_i\}_{i=1}^N$ and vice versa. Some favorable distributions to represent $\mathbf{\Phi}$ are:

1. Gaussian: $\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{M})$

2. Bernouilli/Rademacher: $\phi_{i,j} = \begin{cases} +\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2} \\ -\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{2} \end{cases}$

3. Database-friendly: $\phi_{i,j} = \begin{cases} +\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{6} \\ 0 & \text{with probability } \frac{2}{3} \\ -\frac{1}{\sqrt{M}} & \text{with probability } \frac{1}{6} \end{cases}$

4. Random orthoprojection to $\mathcal{R}^M$.

A Gaussian measurement matrix has an important and useful property: the matrix $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ is also independent and identically distributed (i.i.d) Gaussian regardless of the choice of the sparsifying basis matrix $\mathbf{\Psi}$. Thus, random Gaussian measurements are universal in the sense that $\mathbf{\Phi}$ is incoherent with $\mathbf{\Psi}$ for every possible $\mathbf{\Psi}$ making the reconstruction possible with high probability if $M > cK\log(N)$, with c a small constant.

In order to study the general robustness of the CS measurement matrix, the so-called *Restricted Isometry Property* (RIP) has been proposed by Candès and Tao [24]. For each integer $S = 1, 2, \ldots$, they define the isometry constant δ_S of a matrix $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ as the smallest number such that

$$(1 - \delta_S)\|\mathbf{s}\|_2^2 \leq \|\mathbf{\Theta s}\|_2^2 \leq (1 + \delta_S)\|\mathbf{s}\|_2^2 \tag{2.10}$$

holds for all S-sparse vectors $\mathbf{s}$. A matrix $\mathbf{\Theta}$ is said to obey the RIP of order S if δ_S is not too close to one. When this property holds, $\mathbf{\Theta}$ approximately preserves the Euclidean length of S-sparse signals. An equivalent description of the RIP is to say that all subsets of S columns taken from $\mathbf{\Theta}$ are in fact nearly orthogonal. Therefore, the *mutual coherence* parameter μ represents as well a good measure of robustness

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{N} \cdot \max_{k \leq M, j \leq N} |\langle \boldsymbol{\phi}_k, \boldsymbol{\psi}_j \rangle| \tag{2.11}$$

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) \in [1, \sqrt{N}] \tag{2.12}$$

μ is defined as a measure of the incoherence between the matrices involved in CS and it is proportional to the minimum number of measurements which are needed in order to perfectly reconstruct the sparse vector.

So it is possible to define a universal measurement process, based on projections over a random matrix in which the signal is not sparse. This is possible because even if the projection $\mathbf{\Phi}$ does not have full rank ($M < N$) and loses information in general, it preserves structure and information in sparse signal models with high probability and it is invertible also for sparse models with high probability solving the ill-posed inverse problem.

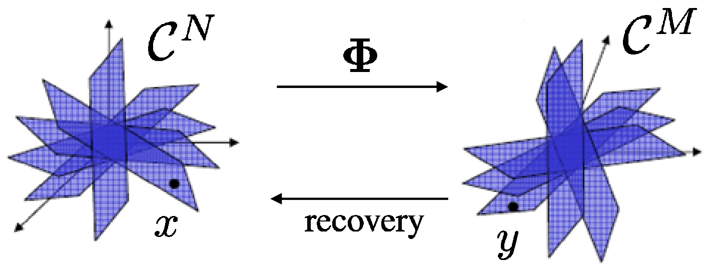

Figure 2.2: Random projection invertible for sparse signals [60].

2.3.2 Signal reconstruction algorithms

The incoherent property provides the theoretical guarantee that a K-sparse or compressible signal can be fully described by the M measurements in $\mathbf{y}$, but it does not tell us how to recover it. The signal reconstruction algorithm must take the measurement $\mathbf{y}$, the random measurement matrix $\mathbf{\Phi}$ (or the random seed that generated it), and the sparsifying basis $\mathbf{\Psi}$ and regenerate the length-N signal $\mathbf{x}$, or equivalently its sparse coefficient vector $\mathbf{s}$.

Define the p-th power of the l_p norm of the vector $\mathbf{s}$ as

$$(\|s\|_p)^p = \sum_{i=1}^{N} |s_i|^p. \tag{2.13}$$

When $p = 0$ we obtain the l_0 norm that counts the number of non-zero entries in $\mathbf{s}$; hence a K-sparse vector has l_0 norm K.

We now provide a brief and incomplete survey of existing approaches. The basic idea behind sparse signal reconstruction is to identify the smallest subset of columns of $\mathbf{\Theta}$, whose linear span contains (approximately) the observations, $\mathbf{y}$. Algorithmic approaches have been proposed for several decades and broadly fall into three categories.

One class of algorithms adopts a greedy search. Examples include, projection pursuit [31], orthogonal matching pursuit (OMP) [32] and tree based matching pursuit (TbMP) [30]. There exist sufficient conditions on the sparseness of $\mathbf{s}$ and singular values of subsets of columns of $\mathbf{\Theta}$ (restricted isometry property [25]) such that the above algorithms stably recover $\mathbf{s}$ with high probability.

A second class of algorithms recursively solves a sequence of iteratively re-weighted linear least-squares (IRLS) problems [33]; recent results [34] for the noiseless case have established sufficient conditions such that the sequence converges to the sparsest solution.

A third class comprises penalized least-squares solutions for $\mathbf{s}$ and has likewise been used for at least four decades. In this class of approaches, parameters are found via the optimization

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{s}\|_p^p \tag{2.14}$$

or, equivalently for some $\epsilon > 0$

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_p \quad \text{s.t.} \quad \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 < \epsilon \tag{2.15}$$

Ridge regression [35] (i.e. Tikhonov regularization) adopts $p = 2$, while basis pursuit [29] and LASSO [36] use $p = 1$. Equation (2.14) has been widely adopted, for example in image reconstruction or radar imaging. With proper choice of the norm, total variation denoising is also an algorithm in this class for $p = 1$ [23]. Elegant results by several authors [25], [27], [28] have demonstrated sufficient conditions on $\mathbf{\Theta}$, and sparsity of $\mathbf{s}$ such that for $p = 1$ the convex problem in (2.15) provides the unique solution to the non-convex task

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \|\mathbf{\Theta s} - \mathbf{y}\|_2^2 < \epsilon. \tag{2.16}$$

These proofs have validated the widespread use of (2.14) and (2.15), providing a deeper understanding, spurring a resurgent interest, and promoting the interpretation as "compressive sampling". As mentioned before, the sufficient conditions on $\mathbf{\Theta}$ are the RIP [25] or a bound on the mutual coherence [28]. A constructive procedure for $\mathbf{\Theta}$ consistent with RIP still remains an open problem [37].

The number of measurements required to capture a class of signals depends on several different factors:

- The sparsity level K of the signals.
- The length N of the signals.
- The coherence between the measurement matrix $\mathbf{\Phi}$ and the sparsity basis $\mathbf{\Psi}$.

To sum up, some important CS properties are as follows:

1. Stable: the acquisition/recovery process is numerically stable.
2. Universal: the same random projections /hardware can be used for any compressible signal class.
3. Asymmetrical: most processing is carried out at the decoder.
4. Encryption of random projections weakly.
5. Democratic: each measurement carries the same amount of information which makes it robust to measurement loss.

2.4 Conclusions

In this chapter we discussed some detection and estimation techniques and we introduced CS as a suitable technology providing a solution to reduce the challenging high sampling frequency rates required by wideband spectrum sensing. In the following chapter, we shall provide a novel acquisition scheme for wideband sensing which obtains an estimate of the PSD from compressive measurements.

Compressive Wideband Sensing

3

In this chapter, we address the problem of estimating the spectrum of the wide-band signal received at the CR sensing receiver using the CS theory introduced in Section 2.3.

In [26], a spectrum sensing scheme based on CS was introduced. The signal received from the licensed system at the CR sensing receiver is sampled, albeit at the Nyquist rate. The autocorrelation of the resulting signal is compressively sampled. An estimate of the spectrum is then obtained using a wavelet edge detector after CS reconstruction, thus determining the spectrum occupancy of the licensed system. This scheme still requires an ADC to operate at Nyquist rate or higher and takes a paradoxical approach to CS. Wide-band ADCs operating at sampling rates of the order of several giga-samples/s are thus a major challenge with such a scheme.

We consider a spectrum sensing scheme based on CS of the wide-band analog signal using an analog-to-information converter (AIC). An AIC directly relates to the idea of sampling at the information rate of the signal. Practical approaches to AIC design have been considered in [51], [50]. An estimate of the original signal spectrum is then made based on CS reconstruction using a wavelet edge detector along the approach in [26]. We evaluate the resulting PSD estimate using the mean squared error (MSE) and the probability of detecting spectrum occupancy, and compare the performance with the scheme in [26]. We note that in [26], CS is done on the autocorrelation of the discrete-time signal obtained by Nyquist-rate sampling. In our approach, CS is directly performed on the wide-band analog signal.

3.1 Preliminaries

Let $x(t)$ be the wide-band analog signal received at the CR sensing receiver. We consider the frequency range of interest to be comprised of P non-overlapping contiguous subbands. The bandwidth and channelization of the subbands need not in general be known to the cognitive radio. For example, it is known that the digital TV licensed system in the UHF-VHF spectrum has a channel occupancy of 8 MHz (in Europe). Each subband may be vacant or used by a PU.

Let the analog signal $x(t)$, $0 \leq t \leq T$, be represented as a finite weighted sum of basis functions (e.g., Fourier) $\psi_i(t)$ as follows

$$x(t) = \sum_{i=1}^{N} s_i \psi_i(t) \tag{3.1}$$

where only a few basis coefficients s_i are much larger than zero due to the sparsity of $x(t)$. In particular, with a discrete-time CS framework, consider the acquisition of an

$N \times 1$ vector $\mathbf{x} = \mathbf{\Psi s}$, where $\mathbf{\Psi}$ is the $N \times N$ sparsity basis matrix and $\mathbf{s}$ an $N \times 1$ vector with $K \ll N$ non-zero (and large enough) entries s_i. It has been shown that $\mathbf{x}$ can be recovered using $M = K\mathcal{O}(\log N)$ non-adaptive linear projection measurements on to an $M \times N$ basis matrix $\mathbf{\Phi}$ that is incoherent with $\mathbf{\Psi}$ [24].

3.2 Compressive spectrum sensing scheme of [26]

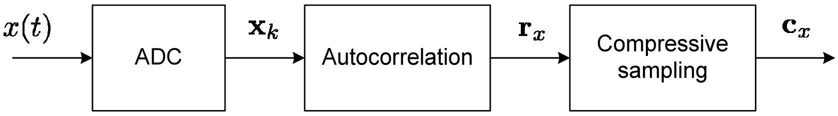

Figure 3.1: CS acquisition in spectrum sensing method of [26].

We now provide a brief overview of the approach of [26]. Figure 3.1 depicts the CS acquisition employed. The basic idea of this approach is to view the entire wide-band spectrum as subbands where subband edges indicate a change in spectrum occupancy. These spectrum edges can be detected using a wavelet-based detector. The CS method is applied to wide-band spectrum sensing as follows. The received signal $x(t)$ is down-converted and sampled at Nyquist rate or higher and the discrete-time signal is stacked in to $N \times 1$ vectors

$$\mathbf{x}_k = [x_{kN} \quad x_{kN+1} \quad \cdots \quad x_{kN+N-1}]^T, \quad k = 0, 1, 2, \ldots \tag{3.2}$$

where ^{T} denotes the transpose operation. We assume the signal to be zero-mean, wide-sense stationary. Denote the autocorrelation at lag j as $r_x(j) = E[x_n x^*_{n-j}]$. In practice, estimates of the autocorrelation are obtained by averaging over several signal segments. Denote the $2N \times 1$ autocorrelation vector of (3.2) as

$$\mathbf{r}_x = [0 \; r_x(-N+1) \; \cdots \; r_x(0) \; \cdots \; r_x(N-1)]^T. \tag{3.3}$$

A wavelet-based smoothing is then performed, followed by taking a Fourier transform to obtain the PSD. Denote the discrete counterparts of these operations by the $2N \times 2N$ matrices $\mathcal{W}$ and $\mathcal{F}$. The derivative of the PSD then gives the edge spectrum. The derivative can be approximated by a first-order difference, given by the $2N \times 2N$ matrix

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Denote $\mathbf{G} = (\mathbf{\Gamma}\mathcal{F}\mathcal{W})^{-1}$. Note that $\mathbf{G}$ represents the transform domain where the autocorrelation vector $\mathbf{r}_x$ has a sparse representation. The $2N \times 1$ discrete component vector $\mathbf{z}_s$ corresponding to the edge spectrum is then related to $\mathbf{r}_x$ by [26]

$$\mathbf{r}_x = \mathbf{G}\mathbf{z}_s \tag{3.4}$$

Compressive sampling is now performed by means of a $2M \times 2N$ compressive matrix $\mathbf{\Phi}_I$, giving rise to the $2M \times 1$ measurement vector $\mathbf{c}_x = \mathbf{\Phi}_I \mathbf{r}_x$. An estimate $\hat{\mathbf{z}}_s^{(1)}$ of the edge spectrum is obtained by solving the CS reconstruction problem:

$$\hat{\mathbf{z}}_s^{(1)} = \arg\min_{\mathbf{z}_s} \|\mathbf{z}_s\|_1 \quad \text{s.t.} \quad \mathbf{c}_x = (\mathbf{\Phi}_I \mathbf{G})\mathbf{z}_s. \tag{3.5}$$

An estimate of the wide-band spectrum can be obtained from $\hat{\mathbf{z}}_s^{(1)} = [\hat{z}_s^{(1)}(1) \;\; \hat{z}_s^{(1)}(2) \;\; \cdots \;\; \hat{z}_s^{(1)}(2N)]^T$ by computing a cumulative sum. The discrete components of the PSD estimate are given by

$$\hat{S}_x^{(1)}(n) = \sum_{k=1}^{n} \hat{z}_s^{(1)}(k). \tag{3.6}$$

It is important to point out that this scheme results in a somewhat paradoxical architecture since sub-Nyquist sampling is achieved by first sampling the wide-band analog signal at Nyquist rate and then applying CS on the autocorrelation vector $\mathbf{r}_x$.

3.3 Compressive spectrum sensing with AIC

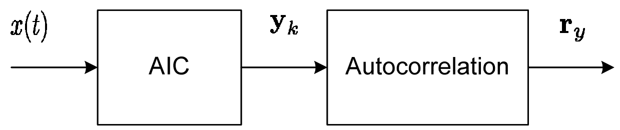

Figure 3.2: CS acquisition in proposed spectrum sensing method.

Figure 3.2 depicts the acquisition under the proposed method. The analog baseband signal $x(t)$ is sampled using an AIC. An AIC may be conceptually viewed as an ADC operating at Nyquist rate, followed by compressive sampling. Denote the $N \times 1$ stacked vector at the output of the ADC by

$$\mathbf{x}_k = [x_{kN} \quad x_{kN+1} \quad \cdots \quad x_{kN+N-1}]^T \quad k = 0, 1, 2 \ldots \tag{3.7}$$

and the $M \times N$ compressive sampling matrix by $\mathbf{\Phi}_A$. The (i, j)-th element of $\mathbf{\Phi}_A$ is given by $\phi_{i,j}$. The output of the AIC denoted by the $M \times 1$ vector

$$\mathbf{y}_k = [y_{kM} \quad y_{kM+1} \quad \cdots \quad y_{kM+M-1}]^T \quad k = 0, 1, 2 \ldots \tag{3.8}$$

is given by

$$\mathbf{y}_k = \mathbf{\Phi}_A \mathbf{x}_k. \tag{3.9}$$

The respective $N \times N$ and $M \times M$ autocorrelation matrices of the compressed signal and the input signal vectors in (3.8) and (3.7) are related as follows

$$\mathbf{R}_y = E[\mathbf{y}_k \mathbf{y}_k^H] = \mathbf{\Phi}_A \mathbf{R}_x \mathbf{\Phi}_A^H \tag{3.10}$$

where ^{H} denotes the Hermitian. The elements of the matrices in (3.10) are given by: $[\mathbf{R}_y]_{ij} = r_y(i - j) = r_y^*(j - i)$, $[\mathbf{R}_x]_{ij} = r_x(i - j) = r_x^*(j - i)$.

Denote the respective $2N \times 1$ and $2M \times 1$ autocorrelation vectors corresponding to (3.7) and (3.8) as follows

$$\mathbf{r}_x = [0 \;\; r_x(-N+1) \;\; \cdots \;\; r_x(0) \;\; \cdots \;\; r_x(N-1)]^T, \tag{3.11}$$

$$\mathbf{r}_y = [0 \;\; r_y(-M+1) \;\; \cdots \;\; r_y(0) \;\; \cdots \;\; r_y(M-1)]^T. \tag{3.12}$$

To pose the CS reconstruction in the form of (3.5), we need to first relate the autocorrelation vectors in (3.11) and (3.12). Note that the components of these vectors lie on the first column and row of the respective autocorrelation matrices. After some matrix algebraic operations, we obtain the following result.

$$\mathbf{r}_y = \mathbf{\Phi}_{II}\mathbf{r}_x \tag{3.13}$$

where $\mathbf{\Phi}_{II}$ is given as

$$\mathbf{\Phi}_{II} = \begin{bmatrix} \overline{\mathbf{\Phi}}_A\mathbf{\Phi}_1 & \overline{\mathbf{\Phi}}_A\mathbf{\Phi}_2 \\ \mathbf{\Phi}_A\mathbf{\Phi}_3 & \mathbf{\Phi}_A\mathbf{\Phi}_4 \end{bmatrix}, \tag{3.14}$$

the $M \times N$ matrix $\overline{\mathbf{\Phi}}_A$ has its (i, j)-th element given by

$$[\overline{\mathbf{\Phi}}_A]_{i,j} = \begin{cases} 0 & i = 1, \; j = 1, \cdots, N, \\ \phi_{M+2-i,j} & i \neq 1, \; j = 1, \cdots, N, \end{cases}$$

and the $N \times N$ matrices $\mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3, \mathbf{\Phi}_4$ are

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & \phi^*_{1,1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \phi^*_{1,1} & \cdots & \phi^*_{1,N-1} \end{bmatrix},$$

$$\mathbf{\Phi}_2 = \begin{bmatrix} \phi^*_{1,1} & \phi^*_{1,2} & \cdots & \phi^*_{1,N} \\ \phi^*_{1,2} & \phi^*_{1,3} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \phi^*_{1,N} & 0 & \cdots & 0 \end{bmatrix},$$

$$\mathbf{\Phi}_3 = \begin{bmatrix} 0 & \cdots & \phi_{1,3} & \phi_{1,2} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \phi_{1,N} \\ 0 & \cdots & \cdots & 0 \end{bmatrix},$$

$$\mathbf{\Phi}_4 = \begin{bmatrix} \phi_{1,1} & 0 & \cdots & 0 \\ \phi_{1,2} & \phi_{1,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \phi_{1,N} & \phi_{1,N-1} & \cdots & \phi_{1,1} \end{bmatrix}.$$

Now using (3.4) and (3.13), we can formulate the CS reconstruction of the edge spectrum as an l_1-norm optimization problem

$$\hat{\mathbf{z}}_s^{(2)} = \arg\min_{\mathbf{z}_s} \|\mathbf{z}_s\|_1 \quad \text{s.t.} \quad \mathbf{r}_y = (\mathbf{\Phi}_{II}\mathbf{G})\mathbf{z}_s \tag{3.15}$$

An estimate of the wide-band spectrum can now be obtained, as done in Section 3.2, from $\hat{\mathbf{z}}_s^{(2)} = [\hat{z}_s^{(2)}(1)\ \hat{z}_s^{(2)}(2)\ \cdots\ \hat{z}_s^{(2)}(2N)]^T$ by computing a cumulative sum. The discrete components of the PSD estimate are given by

$$\hat{S}_x^{(2)}(n) = \sum_{k=1}^{n} \hat{z}_s^{(2)}(k). \tag{3.16}$$

In [24], the mutual coherence parameter μ is defined as a measure of the incoherence between the compressive sampling matrix $\mathbf{\Phi}$ and sparsity basis matrix $\mathbf{\Psi}$ involved in CS,

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{2N} \cdot \max_{1 \le k \le 2M, 1 \le j \le 2N} |\langle \boldsymbol{\phi}_k, \boldsymbol{\psi}_j \rangle|, \tag{3.17}$$

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) \in [1, \sqrt{2N}]$$

where $\boldsymbol{\phi}_k$ and $\boldsymbol{\psi}_j$ are respective columns of $\mathbf{\Phi}$ and $\mathbf{\Psi}$. The proposed scheme incurs a reduced mutual incoherence due to the structure of $\mathbf{\Phi}_{II}$ in (3.14). However this does not have a substantial impact on the performance of spectrum estimation and subsequent detection, as will be shown via simulation results.

3.4 Multiband spectrum detection

We model the detection problem over the subband p as one to choose between hypothesis $\mathcal{H}_{0,p}$ ("0"), which represents the absence of PU signals, and hypothesis $\mathcal{H}_{1,p}$ ("1"), which represents the presence of PU signals. The crucial task of spectrum sensing is to sense the P frequency bands and identify spectral holes for opportunistic use. A convenient description for these signals is the multiband model where the frequency support of a signal resides within several continuous intervals in a wide spectrum but vanishes elsewhere. For simplicity, we assume that the high-layer protocols guarantee that all CRs keep quiet during the detection such that the main spectral power under detection is emitted by the PUs. One measure to compute an optimal detector is to get the maximum p_d for a given p_{fa}. The approach used to compute the threshold that results in a maximum probability of detection for a defined probability of false alarm is called the *Neyman-Pearson theorem* [8]. There exist other measures to compute the performance of a detector, such as Bayesian Approach that minimizes a risk function instead of the p_{fa}, but this requires the knowledge of the a priori probability of the hypothesis, the so called *a priori* distribution. This *a priori* distribution would basically be the probability of PU presence, which generally is not known, as the detection of the PUs' presence is the actual problem.

In the frequency domain, the received signal at each subchannel can be estimated by computing its PSD. The discrete PSD of a wide-sense stationary stochastic process is defined to be the Discrete Fourier Transform (DFT) of its autocorrelation function $r_x(n)$ (Wiener-Khinchin theorem)

$$S_x(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} r_x(n) e^{-j\omega n}, \quad -\pi < \omega \leq \pi, \tag{3.18}$$

or the discrete counterpart

$$S_{x,p}(k) = \sum_{n=0}^{PK-1} r_x(n-N) e^{-j\frac{2\pi}{2N}n(pK+k)} = |H_p(k)|^2 S_{s,p}(k) + V_p(k), \tag{3.19}$$
$$p = 0, 1, \ldots, P-1$$
$$k = 0, 1, \ldots, K-1$$

where $S_{s,p}(k)$ is the PSD of the PU signal occupying subchannel p, $H_p(k)$ stands for the unknown channel frequency response between the corresponding PU transmitter at subchannel p and the CR sensing receiver, $V_p(k)$ is the PSD of the additive complex white Gaussian noise with zero mean and variance σ_v^2, i.e., $v(n) \sim \mathcal{CN}(0, \sigma_v^2)$ and K is the number of samples per band considered in the approximation, holding $2N = PK$. Two architectures to obtain an estimate of the PSD from compressive measurements were proposed in Sections 3.2 and 3.3. Since the radio channel gain is normally considered to

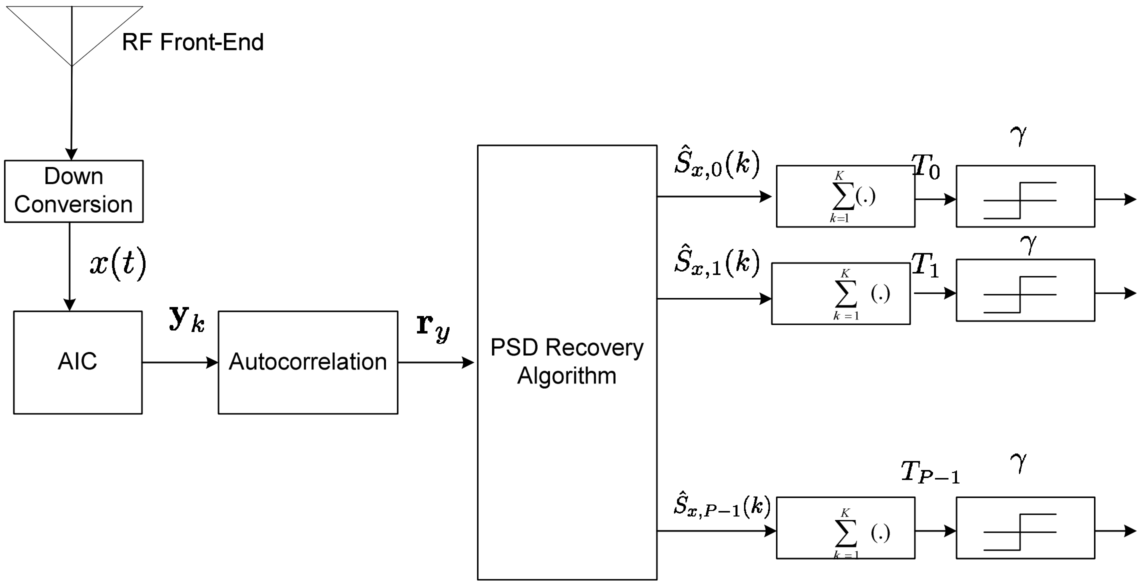

Figure 3.3: A schematic representation of the the independent mutiband detection for wideband spectrum sensing in a CR sensing receiver.

be a random variable, for example the path loss at a particular location is distributed log-normally (normal in dB) when signals are under shadowing effects, the local SNRs in the P sub-channels will take different values. Let us define SNR_p as the ratio of the received signal variance to the noise variance

$$SNR_p = 10 \log_{10}(\|H_p(k)\|_2^2 \sigma_p^2 / \sigma_v^2) \tag{3.20}$$

with σ_p^2 denoting the transmitted energy over subchannel p. To decide whether the p-th subchannel is occupied or not, we test the following binary hypotheses:

$$\mathcal{H}_{0,p} : S_{x,p}(k) = V_p(k), \qquad p = 0, 1, \ldots, P-1$$
$$\mathcal{H}_{1,p} : S_{x,p}(k) = H_p^2(k) S_{s,p}(k) + V_p(k), \qquad k = 0, 1, \ldots, K-1 \tag{3.21}$$
$$\tag{3.22}$$

For each subchannel p, we compute the test statistic as the sum of the received energies, i.e.,

$$T_p = \sum_{k=0}^{K-1} S_{x,p}(k) \quad p = 0, 1, \ldots, P-1. \tag{3.23}$$

The decision rule is given by

$$T_p \underset{\mathcal{H}_{0,p}}{\overset{\mathcal{H}_{1,p}}{\gtrless}} \gamma \qquad p = 0, 1, \ldots, P-1 \tag{3.24}$$

where γ is the corresponding decision threshold. The randomness is coming from the fact that we are estimating the power spectrum density from the finite set of time samples. The threshold γ is found by fixing the p_{fa}. The choice of threshold γ leads to a tradeoff between the probabilities of false alarm and miss $p_m = 1 - p_d$. Specifically, a higher threshold will result in a smaller probability of false alarm but a larger probability of miss, and vice versa. The fundamental tradeoff between p_m and p_{fa} has different implications in the context of dynamic spectrum-sharing. A high p_m would result in missing the presence of a PU with high probability and in turn increases interference to a licensee user. On the other hand, a high p_{fa} would result in low spectrum utilization since false alarms increase the number of missed opportunities (white spaces). Figure 3.3 shows a schematic description of the proposed detection architecture.

3.5 Simulation results

In this section we evaluate the performance of the proposed AIC-based spectrum sensing scheme. We consider, at baseband, a wide frequency band of interest ranging from -40 to 40 MHz, containing $P = 10$ non-overlapping subbands of equal bandwidth of 8 MHz. Each subband is possibly occupied by a licensed system transmission signal that uses OFDM modulation according to the DVB-T standard. Each 8 MHz OFDM symbol has 8192 frequency tones and a cyclic prefix length of 1024. The number of OFDM symbols used for spectrum sensing is 2. The over-sampling[1]factor is 16. The occupancy ratio of the total 80 MHz band is 50%, i.e., 5 out of 10 subbands are occupied by licensed transmission signals and the remaining 5 channels are unoccupied. The received signal is corrupted by additive white Gaussian noise (AWGN) with a

[1]The oversampling is performed commonly in practice after appending the cyclic prefix in order to loosen the requirements on the DAC [59]. When the samples go through the DAC, the spectrum is replicated periodically, and in order to suppress the spectral replicas, a reconstruction filter is required following the DAC. If there are no high frequency bins left empty to create a large gap enough between the spectral replicas, oversampling is required to simplify the design of the reconstruction filter since the roll-off can now be much larger.

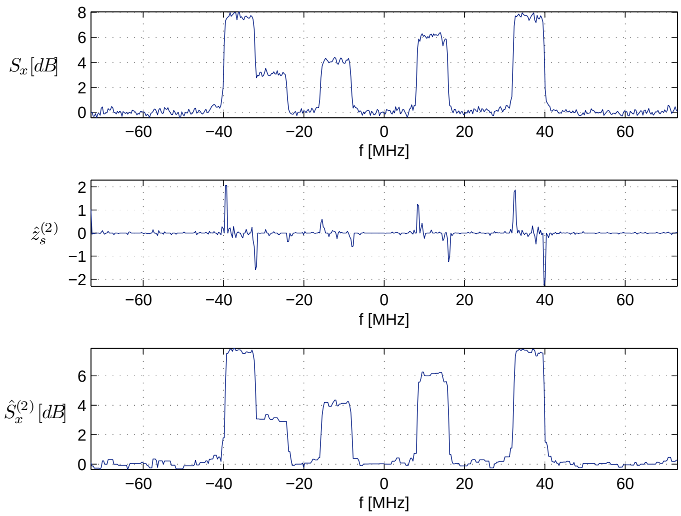

Figure 3.4: Spectrum estimation: (a) Nyquist rate PSD; (b) recovered edge spectrum; (c) recovered PSD from edges

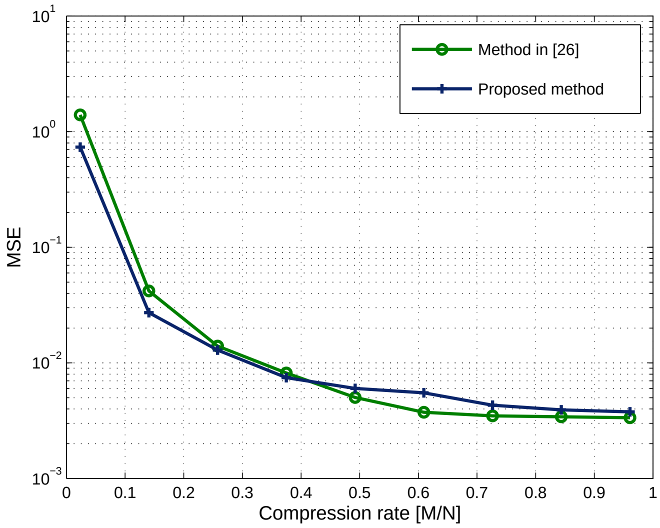

Figure 3.5: MSE performance

variance of $\sigma_n^2 = 1$. The received SNRs of the 5 active channels are 7dB, 0dB, 2dB, 5dB, and 7dB, respectively.

A Gaussian wavelet function is used for smoothing. For compressed sensing, N is 256 and the compression rate M/N is set to vary from 1% to 100%. The entries of the compressive sampling matrix $\mathbf{\Phi}$ are Gaussian distributed with zero mean and variance $1/M$. Seeking to estimate the autocorrelation, we divide the data sequence into segments, thus reducing the variance. The input signal is segmented in time domain

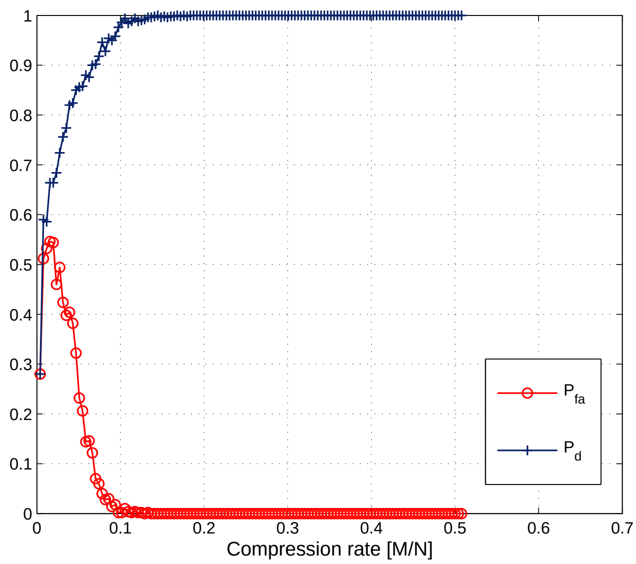

Figure 3.6: Detection performance

and the number of segments is denoted by Q.

The estimated / recovered PSD: Fig. 3.4 shows the estimated PSD based on our proposed approach. The top plot shows the original PSD of the received wide-band signal. The middle plot shows the estimated edge vector $\hat{\mathbf{z}}_s^{(2)}$ in (3.15) using a tree-based Matching Pursuit recovery from the CS measurements with M/N=0.5. The bottom plot shows the recovered PSD $\hat{\mathbf{S}}_x^{(2)}$ vectors whose elements are defined as in (3.16) via a cumulative sum of the estimated edge vector.

MSE performance: We compare the normalized MSE of the estimated PSD of our approach and that of [26]. The normalized MSE is defined as

$$\mathrm{MSE}_i = E\{\frac{\|\hat{\mathbf{S}}_x^{(i)} - \mathbf{S}_x\|_2^2}{\|\mathbf{S}_x\|_2^2}\}, \quad i = 1, 2 \tag{3.25}$$

where $\mathbf{S}_x$ denotes the PSD estimate vector based on the periodogram using the signals sampled at Nyquist rate, $\hat{\mathbf{S}}_x^{(1)}$ the PSD estimate vector based on the approach of [26], and $\hat{\mathbf{S}}_x^{(2)}$ the PSD estimate vector based on our approach. We can see from Fig. 3.5 that for both approaches the signal recovery quality (via tree-based Matching Pursuit) improves as the compression rate M/N increases. The MSE performances of the two approaches are similar, even though a reduced sampling rate is employed in our approach while an ADC operating at Nyquist rate is required in the approach of [26].

Probability of Detection Performance: We evaluate the probability of detection p_d based on the estimated PSD $\hat{\mathbf{S}}_x^{(2)}$. The decision of the presence of a licensed transmission signal in a certain subband is made by an energy detector using the estimated frequency response over that subband. The decision rule is given by

$$T_p \underset{\mathcal{H}_{0,p}}{\overset{\mathcal{H}_{1,p}}{\gtrless}} \gamma, \quad p = 0, 1, \ldots, 9 \tag{3.26}$$

In order to determine the detection threshold, we assume the signal to be ergodic. Therefore, the PSD may be estimated by

$$S_x(b) = \frac{1}{2QN} \sum_{q=0}^{Q-1} |X[q, b]|^2, \quad b = 0, 1, \ldots, 2N-1, \tag{3.27}$$

where

$$X[q, b] = \sum_{l=0}^{2N-1} x[q2N + l] e^{\frac{-j2\pi bl}{2N}}, \tag{3.28}$$

which is the DFT of the non-overlapping sliding sampled signal. The parameter $2N$ is the size of the DFT. In the absence of PUs, $V[u, l],\ l = 0, 1, \ldots, 2N-1$ computed by (3.28) are i.i.d. complex Gaussian random variables with zero-mean and variance $2N\sigma_v^2$. Then, $S_v(b)$ computed by (3.27) are i.i.d. Gamma distributed random variables

$$S_v(b) \sim \Gamma(Q, \zeta) \tag{3.29}$$

where $\zeta = \frac{2N\sigma_v^2}{2NQ}$. The test statistic is given by

$$T_p = \sum_{b=pK}^{(p+1)K-1} S_x(b), \quad p = 0, 1, \ldots, 9, \tag{3.30}$$

which corresponds to the sum of K independent Gamma distributions and

$$T_p \sim \Gamma(KQ, \zeta). \tag{3.31}$$

Then p_{fa} can be evaluated as

$$p_{fa} = Prob\{T_p > \gamma | \mathcal{H}_{0,p}\} = \frac{\Gamma(KQ, \frac{\gamma}{\zeta})}{\Gamma(KQ)} \tag{3.32}$$

where γ denotes the energy threshold, $\Gamma(a, x)$ is the incomplete gamma function given by $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$ and $\Gamma(a)$ is the gamma function. From the Central Limit Theorem, when KQ is sufficiently large, T_p approaches that of a Gaussian distribution

$$\lim_{KQ \to \infty} T_p \ \rightarrow \ N(KQ\zeta, KQ\zeta^2). \tag{3.33}$$

Therefore, the cumulative distribution function of the test statistic for hypothesis $\mathcal{H}_0$ is given by

$$F_{T_p} = \mathrm{Q}(\frac{\gamma - K}{\sqrt{\frac{K}{Q}}}) \tag{3.34}$$

where Q$(\cdot)$ corresponds to the standard normal cdf

$$\mathrm{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \tag{3.35}$$

Then, for a particular p_{fa}, the corresponding threshold γ can be found by

$$p_{fa} = 1 - F_{T_p}(\gamma : \mathcal{H}_{0,p}). \tag{3.36}$$

Finally, after some straightforward calculations, we have

$$\gamma = \sigma_v^2 \sqrt{\frac{K}{Q}} \mathrm{Q}^{-1}(1 - p_{fa}) + K\sigma_v^2. \tag{3.37}$$

The threshold γ is found by fixing p_{fa} to 0.01, and $Q = 288$. The probability of detection p_d is calculated as

$$p_d = Prob\{T_p > \gamma | \mathcal{H}_{1,p}\} = \frac{1}{5} \sum_{p=p_1}^{p_5} Pr\{T_p > \gamma\} \tag{3.38}$$

where $p_i, i = 1, \ldots, 5$ denote the indices of five active subbands. Fig. 3.6 shows p_d versus different values of compression rate M/N under a fixed p_{fa} of 0.01. Note that p_{fa} depends only on the noise variance, thus the threshold can be set regardless of the PU signal level. We observe that p_d reaches 1 when M/N is larger than 0.1. When M/N is smaller than 0.1, p_{fa} is much larger than the designed value of 0.01, which indicates that the CS recovery completely fails due to too few CS measurements.

3.6 Conclusions

We presented a compressive wide-band spectrum sensing scheme wherein an AIC operates on the received analog signal. Spectrum estimation is done based on CS reconstruction using the autocorrelation vector of the resulting compressed signal. The spectrum estimate was used to determine the spectrum occupancy of the licensed system. Performance evaluation using MSE and probability of detection showed that the proposed scheme performs comparably to the scheme in [26]. The loss in incoherence thus does not substantially affect spectrum estimation and spectrum occupancy detection.

4 Distributed Spectrum Sensing

In the previous chapter we proposed an architecture to perform energy based compressive wide-band detection at a local CR sensing receiver. However, if one CR does not see energy in a particular band, it cannot assume that the PU is not present. After all, a SU may suffer multi-path and/or severe shadowing with respect to the PU transmitter. At the same time, its own transmissions may interfere with a primary receiver should it decide to transmit (classical "hidden terminal problem" in wireless networks). To account for possible losses from multi-path, shadowing and local interference, the SU must be significantly more sensitive in detecting than the primary receiver. The presence of multiple radios helps to reduce the effects of severe multi-path at a single radio since they provide multiple independent realizations of related random variables (exploiting spatial diversity). With multiple realizations, the probability that all users see deep fades is extremely low. In essence we wish to make the cognitive radios' spectrum sensing robust to severe or poorly modeled fading environments.

In this chapter we consider the situation in which spectrum sensing is compromised by destructive channel conditions between the PUs and the detecting CRs, which makes it hard to distinguish between an empty spectrum band and a weak signal. We propose a distributed wide-band spectrum sensing scheme that exploits the joint common structure of the received PU signal among CRs to improve the sensing reliability via CS.

Each radio performs spectrum sensing locally, subsequently, coordination among the nodes to share their statistics about the common phenomenon sensed is required. The sharing mechanisms depend on the network architecture. We consider a centralized network, relying on a fusion center (FC) playing the role of a coordinator for exchange of spectrum sensing measurements of local CR sensing receivers. The local sensing measurements can be obtained following any of the acquisition schemes described in Chapter 3. However, along this chapter we assume every node is equipped with the acquisition sensing receiver proposed in Section 3.3.

4.1 Distributed compressive spectrum sensing

Because detection by one CR receiver is subject to missed detection due to channel fading and low SNR, a network of spatially diverse CRs is required. In other words, local spectrum sensing can never surpass its limitation on detecting weak signals. Distributed spectrum decision fusion mitigates the channel fading effect by enabling spatial diversity gain. Diversity combining exploits the fact that independent fading signals have a low probability to be all in a deep fade at the same time. The impact of diversity schemes such as equal gain combining (EGC), selection combining (SC) and switch and stay combining (Dual SSC) over various fading channels is studied in [46]. There are

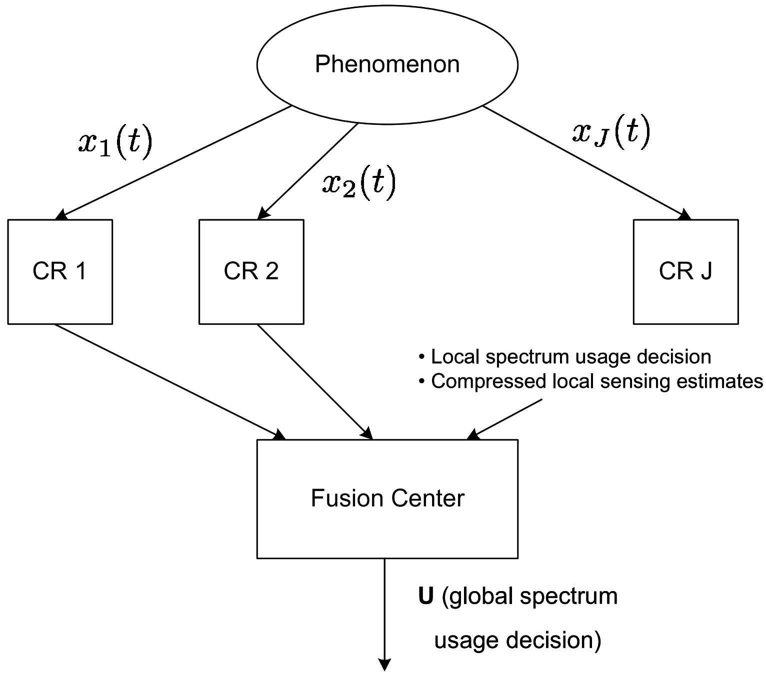

Figure 4.1: Block diagram of a parallel fusion network.

previous studies on spectrum sensing in CR networks with focus on cooperation among multiple CRs [13] [44] [45], via distributed detection approaches [43]. However, they are limited to the detection of signals on a single frequency band. A cooperative wide-band spectrum sensing scheme was proposed in [47]. To provide reliable spectrum sensing at affordable complexity, we present a distributed compressed edge sensing framework for wide-band communication networks.

In this section, we model the CR system with a standard parallel fusion network with a total of J sensing receivers randomly deployed, as shown in Fig. 4.1. We assume that the nodes do not communicate with each other and that there is no feedback from the FC to any node. In this model, each CR obtains some relevant information on the spectrum occupancy and they rely on a FC to make collaborative decisions with improved sensing quality. Let

$$x_j(t) \quad j = 1, 2, \ldots, J, \tag{4.1}$$

be the analog wideband signal received at the j-th CR sensing receiver. We propose two architectures:

1. *Independent CS* (Fig. 4.2): In short, each CR sensing receiver makes a binary decision based on its local observation and then forwards its per-band decision to the FC. After gathering all decision vectors, the FC generates a global spectrum usage decision by fusing the local decisions. In this approach we seek to minimize the communication overhead. Subsequently, users only share their final decisions rather than the raw data set.

2. *Joint CS* (Fig. 4.3): Briefly, each CR node sends the compressed local sensing estimates $\mathbf{r}_{y,j}$ to the fusion center which jointly estimates the J replicas and then gen-

erates a global spectrum usage decision. In this approach we seek to reduce complexity at the nodes.

4.1.1 Independent CS

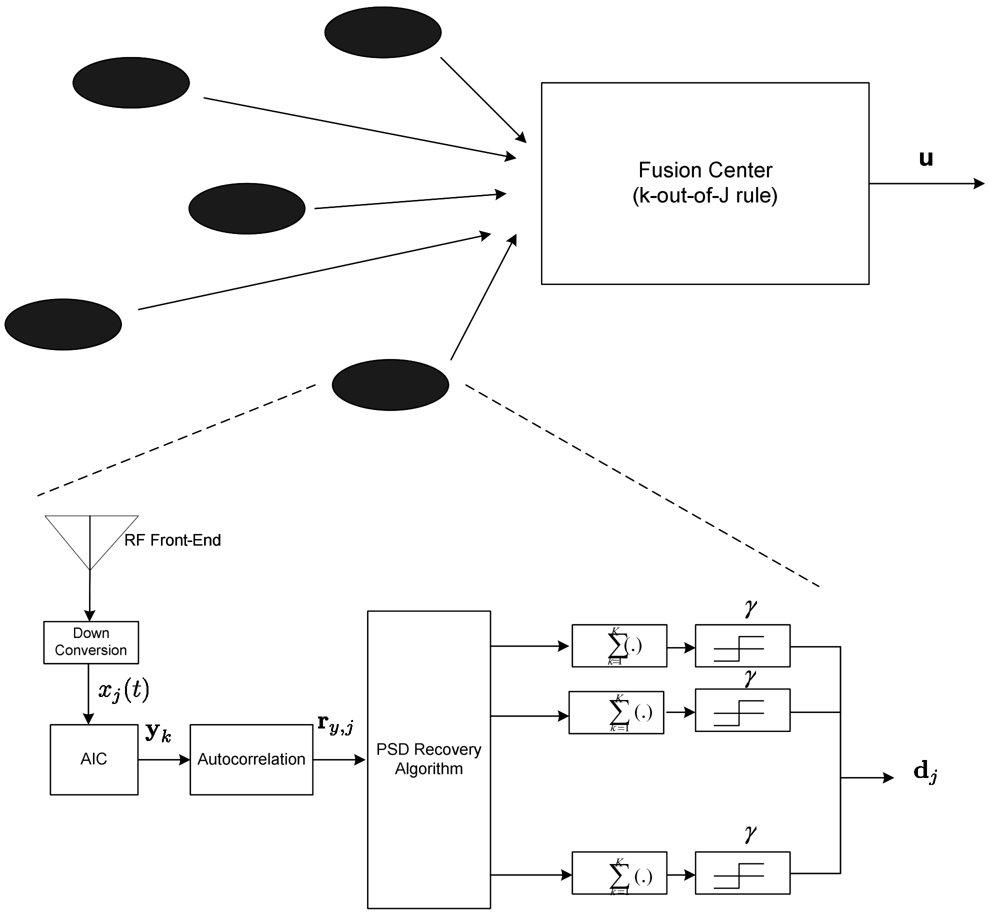

Figure 4.2: Block diagram of the architecture described in Section 4.1.1 (Independent CS).

Fig. 4.2 depicts the sensing architecture under the Independent CS method. The procedure of this scheme is described as follows:

1. Denote $\mathbf{d}_j$ as the $P \times 1$ decision vector obtained at the CR sensing receiver j

$$\mathbf{d}_j = [d_{1,j}\ d_{2,j}\ \cdots\ d_{P,j}]^T \quad j = 1, 2, \ldots, J. \tag{4.2}$$

   Every CR j performs wideband spectrum detection independently obtaining a binary decision vector $\mathbf{d}_j$ (following the NP energy detector introduced in Section 3.4) being

$$d_{p,j} = \begin{cases} 1, & \mathcal{H}_{1,p} \\ 0, & \mathcal{H}_{0,p} \end{cases} \tag{4.3}$$

2. All of the CRs forward their binary decisions vector to a common FC.

3. Denote the $P \times 1$ global spectrum usage decision vector as

$$\mathbf{u} = [u_1 \ u_2 \ \cdots \ u_P]^T. \tag{4.4}$$

The FC combines those binary decisions and makes a final decision $\mathbf{u}$ given by

$$u_p = \begin{cases} 1, & \mathcal{H}_{1,p} \\ 0, & \mathcal{H}_{0,p} \end{cases} \tag{4.5}$$

to infer the absence or presence of the PU in the P frequency bands.

For simplicity we assume that all J cognitive users experience independent and identically distributed fading/shadowing with the same average SNR. A fundamental result in distributed binary hypothesis testing is that when nodes are conditionally independent (as in our case), the optimal decision rule for individual nodes is the likelihood ratio test (LRT) [43]. However, optimum individual thresholds are not necessarily equal and it is generally hard to derive them. We assume that all users employ a NP energy-detector and use the same decision rule (i.e. same threshold). While these assumptions render our scheme sub-optimum, they facilitate analysis as well as practical implementation. As long as the threshold is known, the probability of false alarm at each CR sensing receiver is known, which can be derived from the probability density function (PDF) of the noise. However, at each node, it is very difficult to calculate the probability of detection p_d, since it is determined by each CR sensing receiver's channel with the PU. Without the knowledge of p_d, the FC is forced to treat detections from every SU equally. At the common receiver, all 1-bit decisions are fused together according to the following logic k-out-of-J rule

$$u_p = \begin{cases} 1, & \text{if } \sum_{j=1}^{J} d_{p,j} \geq k \\ 0, & \text{if } \sum_{j=1}^{J} d_{p,j} < k \end{cases} \tag{4.6}$$

This means that the FC adopts hypothesis $\mathcal{H}_{1,p}$ (presence of a PU) as the true hypothesis when at least k CRs favore that hypothesis. Under hypothesis $\mathcal{H}_{1,p}$ (PU presence), the probability of detection at CR sensing receiver j is a function of its receiver's SNR. Under hypothesis $\mathcal{H}_{0,p}$ (PU absence), each SU has the same probability of false alarm, and we denote it as p_{fa}. At the FC, the global probability of false alarm P_{FA} is given by

$$P_{FA} = Prob\{u_p > k|\mathcal{H}_0\} = \sum_{i=k}^{J} c_i^J p_{fa}^i (1 - p_{fa})^{J-i} \tag{4.7}$$

and the global probability of detection P_D, assuming that all decisions are independent from each other under hypothesis $\mathcal{H}_1$

$$P_D = Prob\{u_p > k|\mathcal{H}_1\} = \sum_{i=k}^{J} c_i^J p_d^i (1 - p_d)^{J-i} \tag{4.8}$$

where

$$c_i^J = \frac{J!}{i!(J-i)!} \tag{4.9}$$

For $k = 1$, the fusion rule reduces to an OR fusion rule, while for $k = J$ it becomes an AND fusion rule. For a specified value of P_{FA}, there is an integer k that maximizes the P_D. In [42] it is shown that when minimizing the total probability of error when p_{fa} and p_m have the same order, the optimal choice of k is $J/2$, the OR rule becomes optimal when $p_{fa} \leq p_m^{J-1}$ (large detection threshold γ) and the AND rule is optimal when $p_m \ll p_{fa}$ (small detection threshold γ).

4.1.1.1 Decentralized Independent CS

It is worth mentioning that when seeking for completely decentralized algorithms a "consensus" or "agreement computation" for networked systems can be followed, i.e., in the absence of a fusion center. These systems are more robust than the centralized ones since the network cannot be compromised just by eliminating the FC (node failure). Also, since all nodes asymptotically compute the consensus value there is an added layer of robustness. In such ad hoc CR systems, each node is a SU equipped with a CR acting as both a sensing terminal and a FC.

In order to solve the decentralized distributed averaging problem, let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be an undirected connected graph with node set $\mathcal{N} = \{1, \ldots, J\}$ and edge set $\mathcal{E}$, where each edge $(j, k) \in \mathcal{E}$ is an unordered pair of distinct nodes. Let $\mathbf{c}_j(0) = \mathbf{d}_j$ be a real vector associated with CR node j at time $t = 0$. The average consensus problem is to compute the average $\frac{1}{J}\sum_{j=1}^{J} \mathbf{c}_j(0)$ at every node, via local communication with its neighbors $\mathcal{N}_j = \{k|(j, k) \in \mathcal{E}\}$.

Gossip algorithms [48] provide an intuitive solution by computing a sequence of pairwise averages. In each round, one node is chosen randomly, and it chooses one of its neighbors randomly. Both nodes compute the average of their values and replace their own value with this average. By iterating this pairwise averaging process, the estimates of all nodes converge to the global average under suitable conditions on the graph topology.

Alternatively, each node may update its local variable by adding a weighted sum of the local discrepancies, i.e., the differences between neighboring node values and its own

$$\mathbf{c}_j(t+1) = \mathbf{c}_j(t) + \sum_{k \in \mathcal{N}_j} \mathbf{W}_{jk}(\mathbf{c}_k(t) - \mathbf{c}_j(t)) \qquad j = 1, \ldots, J;\ t = 0, 1, \ldots, \tag{4.10}$$

where $\mathbf{W}_{jk}$ is a weight associated with the edge (j, k). These weights are algorithm parameters. Since we associate weights with undirected edges, we have $\mathbf{W}_{jk} = \mathbf{W}_{kj}$. In [49], the problem of finding the edge weights that result in the least mean square deviation in steady state was considered. With properly designed weights, we can guarantee that

$$\lim_{t \to \infty} \mathbf{c}_j(t) = \frac{1}{J}\sum_{k=1}^{J} \mathbf{c}_k(0) = \frac{1}{J}\sum_{k=1}^{J} \mathbf{d}_k, \quad j = 1, \ldots, J. \tag{4.11}$$

Thus, through local one-hop communications, each CR obtains the averaged statistic of the entire multi-hop network. Subsequently, each CR can make the fusion decision straightforwardly by comparing $\mathbf{c}_j(t)$ with a threshold

$$\gamma_{ac} \in [\frac{1}{J}, 1] \tag{4.12}$$

at a sufficiently large t.

The threshold is chosen to reflect how conservative the network is in protecting the PUs. In the most conservative case, detection is declared as long as there exists a single CR which locally declares detection, corresponding to $\gamma_{ac} = \frac{1}{J}$ (OR fusion rule).

The fundamental issue in average consensus is the number of iterations it takes to converge to a sufficiently accurate estimate.

4.1.2 Joint CS

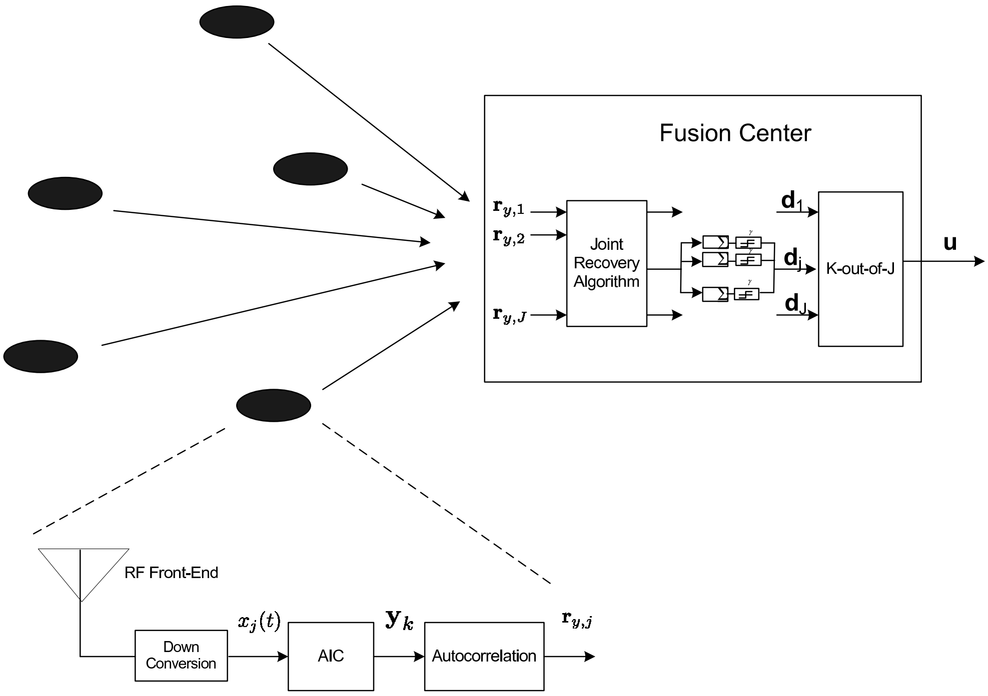

Figure 4.3: Block diagram of the architecture described in Section 4.1.2 (Joint CS).

In [38] the compressed sensing theory was extended to take into account the joint sparsity of a signal ensemble giving rise to the Distributed Compressive Sensing (DCS) framework based on the joint structure of an ensemble of signals in the case they observe presumably related phenomena. In a typical DCS scenario, a number of CR sensing receivers measure signals that are each individually sparse in some basis and also correlated from node to node. Next we present a model for jointly sparse signals which exploits both the intra- and inter-signal correlation structure that allows for joint recovery at the FC from the raw data of the CR nodes. Fig. 4.3 depicts the detection architecture under the joint CS method. The procedure of this scheme is described as follows:

1. Every CR sensing receiver j obtains compressive sensing measurements $\mathbf{r}_{y,j}$ independently which are then forwarded to the FC.

2. The FC jointly reconstructs the J received PSD's $\hat{\mathbf{S}}_{x,j},\ j = 1, \dots, J$, and then makes JP binary decisions $d_{p,j}$.

3. Those binary decisions are combined and a final decision $\mathbf{u}_p$ is made to infer the absence or presence of the PU in the observed frequency band p following the k-out-of-J rule defined in (4.6).

The multiple CR nodes are acquiring the same signal but through a different channel (phase shifts and attenuations caused by signal propagation). Since all CRs are measuring the same phenomenon, all signals are constructed from the same edge-domain sparse set of basis vectors, but with different coefficients.

In order to show more explicitly the above mentioned relationship among the recovered edge spectrum $\mathbf{z}_{s,j}$ in the CR local nodes, let's suppose as done in [41] that there are I PUs emitting spectral power at I of the total P subchannels during the detection interval, whose transmitted signals are denoted by $s_i(t),\ i = 1, \dots, I$. After propagating through a wireless fading channel, the signal $s_i(t)$ reaches the j-th CR sensing receiver in the form $h_{ij}(t) \star s_i(t)$, where $\star$ denotes the convolution operator and $h_{ij}(t)$ is the channel impulse response between the i-th PU and the j-th CR sensing receiver. We assume that the channel is slowly varying such that the channel frequency response remains constant during a detection interval. The received signal at CR j is thus given by

$$x_j(t) = \sum_{i=1}^{I} h_{ij}(t) \star s_i(t) + v_j(t) \tag{4.13}$$

where the ambient noise $v_j(t)$ is white Gaussian with zero mean and PSD σ_v^2. To reflect the discretized signal response on the P sub-channels, we transform equation (4.13) into its discretized PSD-domain. Consequently, following (3.19) we have

$$\mathbf{S}_{x,j} = \mathbf{H}_j \mathbf{H}_j^* \mathbf{S}_s + \mathbf{V}_j \quad j = 1, 2, \dots, J \tag{4.14}$$

Let $\mathbf{z}_s$ denote the $2N \times 1$ edge spectrum of $\mathbf{S}_s$ given by

$$\mathbf{S}_s = \mathcal{F}\mathbf{G}\mathbf{z}_s. \tag{4.15}$$

When channel state information (CSI) is available at each CR receiver j, $\mathbf{z}_s$ may be estimated at every CR sensing receiver. In [41] a computationally distributed fusion technique based on CS is proposed based on a stochastic approximation when CSI is available at the CR sensing receivers. This technique may be directly applied to our framework. However, acquistion of the CSI, may be extremely costly or even impossible in a multiuser wireless system with unknown PU locations and transmission parameters. Consequently, we estimate the faded edge spectrum at the j-th CR sensing receiver $\mathbf{z}_{s,j}$ which is related to $\mathbf{z}_s$ as

$$\mathbf{z}_{s,j} = \mathbf{H}_j \mathbf{H}_j^* \mathbf{z}_s \tag{4.16}$$

Hence, all signals share the same common sparse support (positions of edges) which we denote as Ω. The common structure among the signals may be exploited to perform an improved joint recovery of the set at the FC.

4.1.2.1 Joint recovery algorithm

The common sparse profile allows for a fast algorithm to recover all of the signals jointly. A popular search technique for finding the sparse solution is based on a suboptimal forward search through the dictionary. These algorithms, termed Matching Pursuit (MP), proceed by sequentially adding vectors to a set which will be used to represent the signal. After convergence we obtain an expansion of the measurement vector on a subset of the dictionary basis vectors. To obtain the expansion coefficients in the sparse basis, we then reverse it by least squares.

We write $\mathbf{\Theta} = \mathbf{\Phi}_{II}\mathbf{G}$ in terms of its columns

$$\mathbf{\Theta} = [\theta_{m,n}] = [\boldsymbol{\theta}_1 \ \boldsymbol{\theta}_2 \ \cdots \ \boldsymbol{\theta}_{2N}] \quad \begin{aligned} m &= 1, \ldots, 2M \\ n &= 1, \ldots, 2N \end{aligned} \tag{4.17}$$

Denote Ω as the index set of all columns of the matrix $\mathbf{\Theta}$

$$\Omega \subset \{1, 2, \ldots, 2N\} \tag{4.18}$$

and Ω_k as the index set of all columns that are selected from the beginning up to step k. At the k-th iteration step, the algorithm selects a new column of the common sparse support via a two-stage selection process. In the first stage, the inner products of the current residual are examined with the remaining atoms for all the signals. Based on these inner products, the search is narrowed down to a small set of potential candidates. In the second stage, the atom that provides the maximal average reduction of the residual is selected. The procedure of our algorithm is described as follows:

$\boxed{\text{I.}}$ INPUT:

- A common dictionary $2M \times 2N$ matrix $\mathbf{\Theta}$.
- A $2M \times J$ data matrix $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_J]$ corresponding to the acquired data from the J CR sensing receivers.
- ρ (optional) maximum number of iterations.
- ξ (optional) threshold for convergence.

$\boxed{\text{II.}}$ OUTPUT:

- A $2N \times J$ reconstruction matrix $\mathbf{Z}_s = [\mathbf{z}_{s,1} \ \mathbf{z}_{s,2} \ \cdots \ \mathbf{z}_{s,J}]$ solving

$$\mathbf{\Theta}\mathbf{z}_{s,j} = \mathbf{r}_{y,j} \quad j = 1, 2, \ldots, J, \tag{4.19}$$

with all $\mathbf{z}_{s,j}$ sharing the same sparse support.

$\boxed{\text{III.}}$ PROCEDURE:

1. Denote $\mathbf{R}^{(k)}$ as the $2M \times J$ residual matrix obtained at the k-th iteration written in terms of its columns

$$\mathbf{R}^{(k)} = [\mathbf{r}_1^{(k)} \ \mathbf{r}_2^{(k)} \ \dots \ \mathbf{r}_J^{(k)}]. \tag{4.20}$$

   Initialize the residual $\mathbf{R}^{(0)} = \mathbf{Y}$, the index set $\Omega^{(0)} = 0$, and the counter $k = 1$. Create the index set Ω of all columns in $\boldsymbol{\Theta}$

$$\Omega = \{1, 2 \dots, 2N\}. \tag{4.21}$$

2. Denote $\mathbf{c}_n^{(k)}$ as the $J \times 1$ correlation vector between the residuals and the n-th dictionary element

$$\mathbf{c}_n^{(k)} = [c_{n,1}^{(k)} \ c_{n,2}^{(k)} \ \dots \ c_{n,J}^{(k)}]^T, \tag{4.22}$$

   where

$$\begin{aligned} \mathbf{c}_n^{(k)} = \langle (\mathbf{R}^{(k-1)})^T, \boldsymbol{\theta}_n \rangle \quad & n \in \Omega \backslash \Omega^{(k-1)} \\ & |\Omega \backslash \Omega^{(k-1)}| = 2N - (k-1). \end{aligned} \tag{4.23}$$

   Denote $\bar{\mathbf{c}}^{(k)}$ as the $(2N - (k-1)) \times 1$ average correlation over the CR sensing receiver's vector given by

$$\bar{\mathbf{c}}^{(k)} = [\bar{c}_1^{(k)} \ \dots \ \bar{c}_n^{(k)} \ \dots \ \bar{c}_{2N-(k-1)}^{(k)}]^T, \tag{4.24}$$

   where

$$\bar{c}_n^{(k)} = \frac{1}{J} \sum_{j=1}^{J} c_{n,j}^{(k)} \quad n \in \Omega \backslash \Omega^{(k-1)}. \tag{4.25}$$

   We find the columns of $\boldsymbol{\Theta}$ which have larger common inner products with the residual.

$$c^{*(k)} = \max\{\bar{\mathbf{c}}^{(k)}\} \tag{4.26}$$

$$\Xi^{(k)} = \{n : \bar{c}_n^{(k)} \geq \alpha c^{*(k)}\} \tag{4.27}$$

   with α denoting the narrowing down factor seleceted from 0 to 1.

3. Search among the candidate set $\Xi^{(k)}$ for the item that maximizes the reduction of average residual

$$n^{(k)} = \underset{n \in \Xi^{(k)}}{\operatorname{argmax}} \sum_{j=1}^{J} \|\mathbf{r}_j^{(k)} - \mathbf{P}_{span\{\boldsymbol{\theta}_l \cup \boldsymbol{\theta}_n : l \in \Omega^{(k-1)}\}} \mathbf{r}_j^{(k)}\|_2. \tag{4.28}$$

4. Set $\Omega^{(k)} = \Omega^{(k-1)} \cup n^{(k)}$.
5. Update the residual: $\mathbf{R}^{(k)} = \mathbf{Y} - \mathbf{P}_{span\{\boldsymbol{\theta}_l : l \in \Omega^{(k)}\}} \mathbf{Y}$, where the orthogonal projector onto the span of the selected columns from $\boldsymbol{\Theta}$ is given by

$$\mathbf{P}_{span\{\boldsymbol{\theta}_l : l \in \Omega^{(k)}\}} = \boldsymbol{\Theta}^{(k)} \boldsymbol{\Theta}^{(k)\dagger} \tag{4.29}$$

with

$$[\mathbf{\Theta}^{(k)}]_{m,n} = \theta_{m,n} \quad n \in \Omega^{(k)} \tag{4.30}$$
$$m = 1, 2, \ldots, 2M$$

and $\mathbf{\Theta}^{\dagger}$ denoting the Moore-Penrose pseudo-inverse

$$\mathbf{\Theta}^{\dagger} = (\mathbf{\Theta}^{H}\mathbf{\Theta})^{-1}\mathbf{\Theta}^{H}. \tag{4.31}$$

6. Compare $\min \|\mathbf{r}_j^{(k)}\|_2^2,\ j = 1, 2, \ldots, J$ with a preselected limit ξ, and compare the number of selected items in $\Omega^{(k)}$ with a preselected limit ρ. If these limits are not reached then increase k by one and return to step 2.
7. Locations of nonzero coefficients of $\mathbf{Z}_s$ are listed in $\Omega^{(k)}$. The values of those coefficients are in the expansion

$$\mathbf{P}_{span\{\boldsymbol{\theta}_l : l \in \Omega^{(k)}\}}\mathbf{y}_j = \sum_{l \in \Omega^{(k)}} z_{s,j}(l)\boldsymbol{\theta}_l \rightarrow \mathbf{z}_{s,j} = (\mathbf{\Theta}^{(k)})^{\dagger}\mathbf{y}_j. \tag{4.32}$$

Similar algorithms have been considered by different authors in the area of *simultaneous sparse approximation* [38], [40]. We shall compare the performance of our proposed algorithm against theirs in future work.

4.2 Simulation results

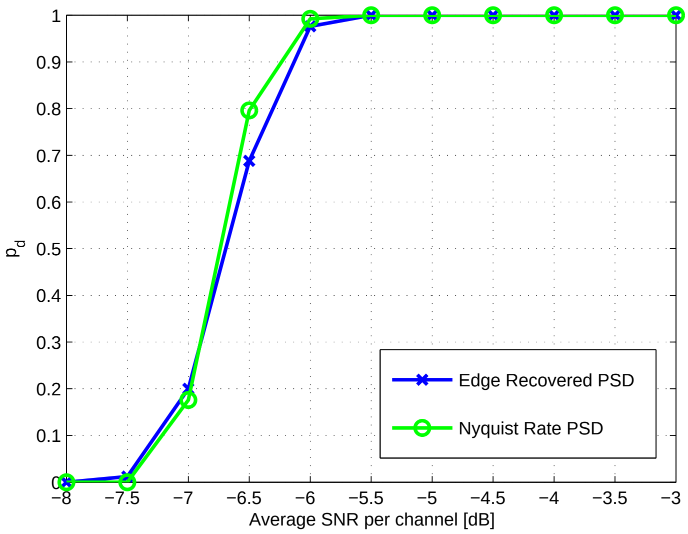

Figure 4.4: Detection performance for fixed average SNR.

In this section we evaluate the performance of the proposed AIC-based distributed spectrum sensing scheme. We consider the simulation scenario described in Section 3.5.

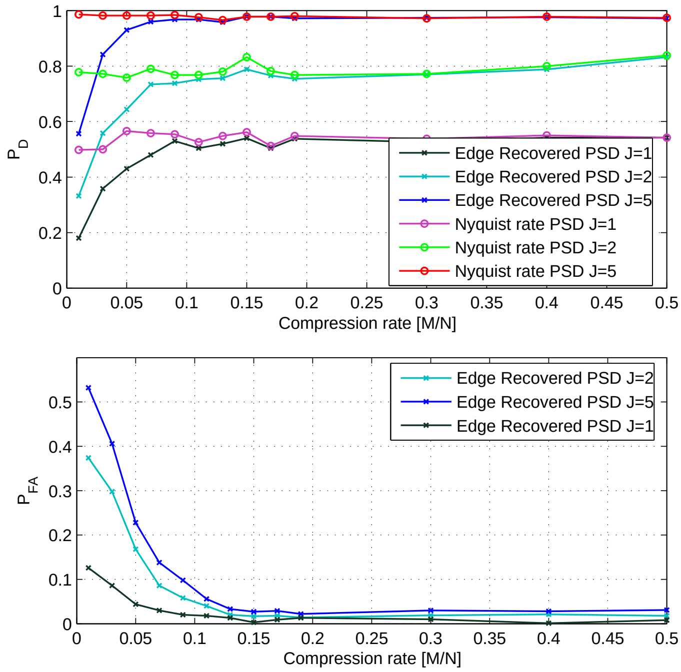

Figure 4.5: Independent Detection performance under low SNR conditions.

Three quantities are varied in the experiments: the SNR per band, the number of CRs J and the compression rate M/N which is set to vary from 1% to 100%. In a Monte Carlo simulator, 500 trials are run. In each trial, a different realization of the measurement matrix $\mathbf{\Phi}$, the sparse vector $\mathbf{z}_{s,j}$, and the noise vectors are used.

Probability of Detection Performance: We evaluate the local probability of detection p_d based on the estimated PSD $\hat{\mathbf{S}}_{x,j}$ and the global probability of detection P_D. The local energy collected in the PSD domain is denoted by T_p and it is defined as in Section 3.5. The local threshold γ is again found by fixing p_{fa} to 0.01 and $Q = 288$. The local probability of detection p_d is calculated as

$$p_d = \frac{1}{5} \sum_{p=p_1}^{p_5} Pr\{T_p > \gamma\} \tag{4.33}$$

where $p_i, i = 1, \dots, 5$ denote the indices of five active subbands. The global probability of detection P_D is computed according to the OR fusion rule.

In Fig. 4.4 we show the local probability of detection p_d against different values of SNR under a fixed p_{fa} of 0.01. The received SNR is held fixed in each set of trials along the different PUs. We observe that there is a SNR_{wall} of -8 dB below which no PU presence is detected.

In Fig. 4.5 and Fig. 4.6 we show the global probability of detection P_D and the false alarm rate P_{FA} when J is held constant while compression rate is set to vary from

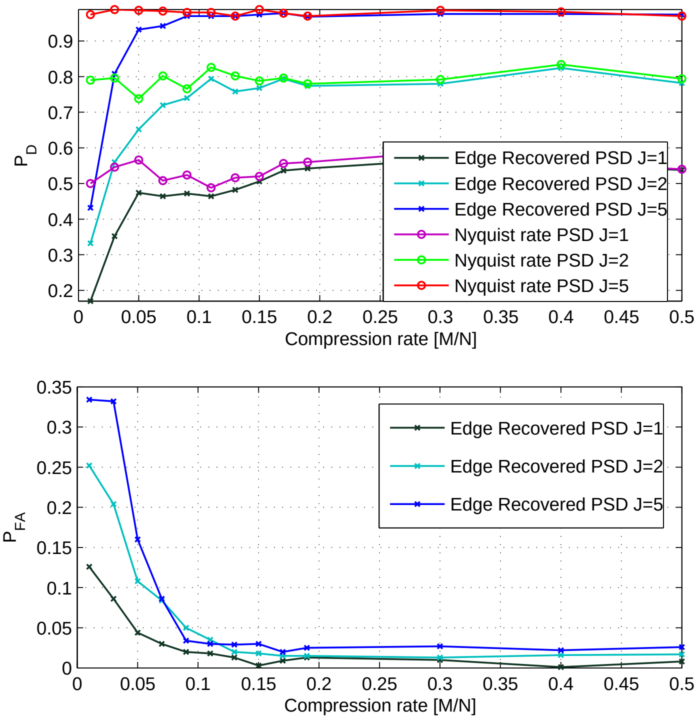

Figure 4.6: Joint Detection performance under low SNR conditions.

1% to 100% and SNR_p also is set to vary from -10 to 0 dB in each trial and from user to user as discussed in Section 4.1.1 and Section 4.1.2, respectively. Note that below $-5.5dB$ local sensing is not sufficient to determine all PUs. The asymptotic bound of the OR fusion probability is determined by the local probability of detection p_d and the number of CRs involved

$$P_D = 1 - (1 - p_d)^J \tag{4.34}$$

$$P_{FA} = 1 - (1 - p_{fa})^J. \tag{4.35}$$

In Fig. 4.7 and Fig. 4.8 we carry out the same experiment as in Fig. 4.5 and Fig. 4.6 with SNR_p varying from 0 to 5 dB.

Spectrum sensing sensitivity requirements are set by the minimum detectable signal strength of sensing SNR:

- In the high noise case ($SNR_p < -5.5$ dB), one CR is not able to identify correctly all the PUs and distributed sensing is required. We note that using a value of $J > 1$ results in an improvement in performance (spatial diversity gain).
- In the low noise case ($SNR_p > -5.5$ dB), the achieved gain when using a distributed network of CRs is highly reduced. For instance, we are able to find the

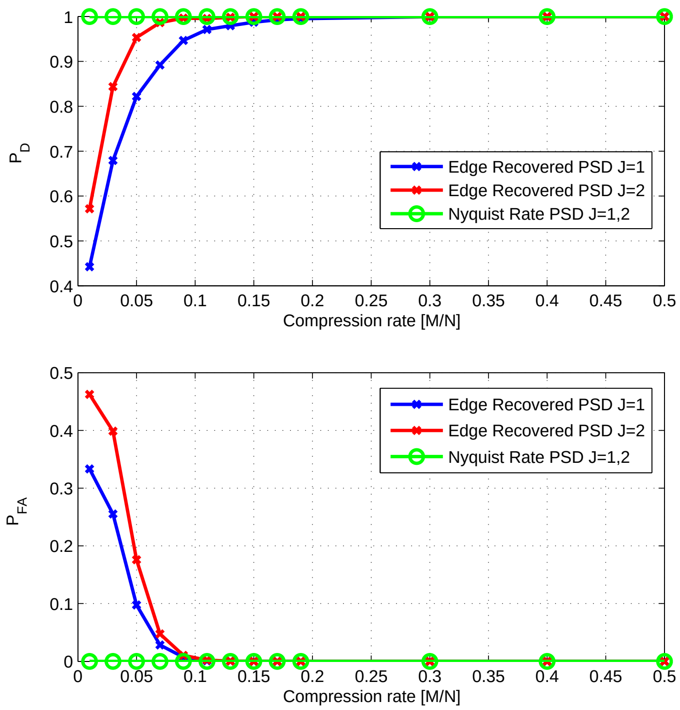

Figure 4.7: Joint Detection performance under high SNR conditions.

correct occupancy with $J = 1$.

- In both high and low noise case, when M/N is smaller than 0.2, p_{fa} is much larger than the designed value of 0.01, which indicates that the CS recovery completely fails due to too few CS measurements.

Finally, we note that the global probability of detection P_D is very similar between the independent distributed scenario and the joint distributed one. This can be explained by the fact that even if the joint recovery algorithms exploits the common sparse support in the high noise case, the recovered signal amplitude will depend on the received SNR which in high noise case turns to be below the noise level leading in both cases to a missed detection.

4.3 Conclusions

We presented a distributed compressive wide-band spectrum sensing architecture based on the compressive acquisition scheme proposed in Chapter 3. We motivated the distributed network of CRs by the need for an additional layer of robustness to fight against unfavorable channel conditions. In order to coordinate the nodes, we assumed the existence of a FC and depending on how the tasks involved in spectrum sensing are

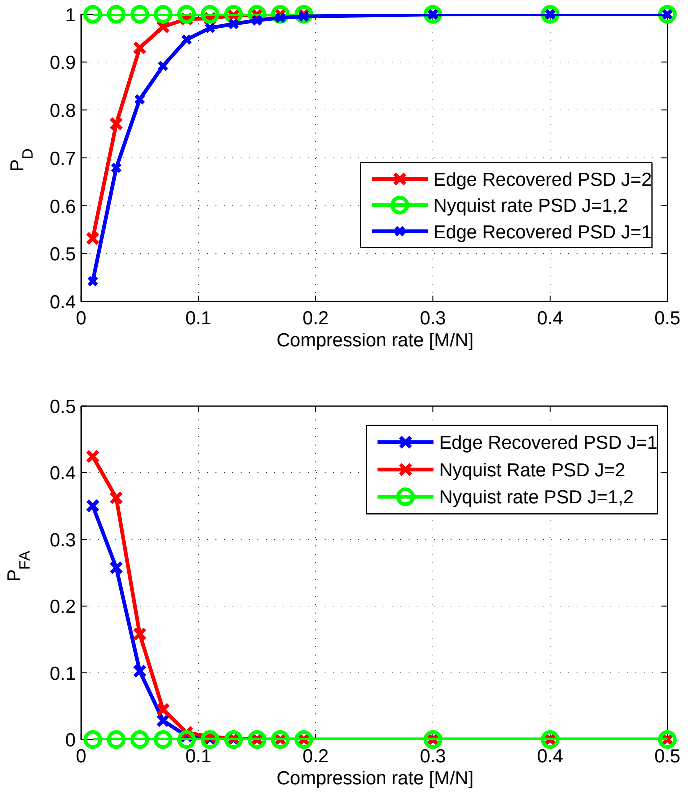

Figure 4.8: Independent Detection performance under high SNR conditions.

shared we proposed two architectures: independent CS and joint CS. Performance evaluation using local probability of detection and global probability of detection showed that both the proposed architectures perform comparably well in destructive and non destructive channel conditions even if no CSI is available.

Practical Performance Issues 5

In this chapter, we address some major practical performance issues related with the proposed architectures. We start with a description of the influence of the main parameters governing the scheme performance. Second we discuss the practical limitations of our spectrum detector block, i.e., noise variance assumed constant and known, local channels of the distributed network assumed i.i.d., ... Then we describe the main architectures which have already been proposed in the literature to implement the AIC and we finally conclude the chapter with a low complexity distributed compressive recovery algorithm to perform detection without going into the intermediate stage of estimating the PSD.

5.1 Degrees of freedom to improve signal detection

All simulations along the thesis have been obtained by fixing the parameters whose influence we now discuss:

- *Spectral estimation block size* (N): this parameter is related to the bandwidth resolution, referred to as the ability to discriminate spectral features.
- *Number of spectral averages* (Q): by averaging we reduce the variance of the PSD estimator. Note that obtaining a PSD estimator with the lowest possible variance is crucial for the compressive edge detector.
- *Measurement matrix* ($\mathbf{\Phi}$): in Section 2.3.1, we have introduced several possible measurement matrices. The main requirement of the measurement matrix is to provide incoherency with the sparsifying matrix. The bigger the incoherency is, the less measurements are needed by the compressive edge optimization algorithm to converge. Owing to the universal incoherency of the gaussian matrix, we selected it to run the experiments all along the thesis.
- *Number of compressive measurements* (M): the applied compression in the scheme influences the error in the PSD estimate. It has been shown from simulations that a compression rate of 10% incurs a very similar performance as the full Nyquist rate signal.
- *Wavelet smoothing function* ($\mathcal{W}$): the matrix $\mathcal{W}$ introduced in Section 3.1 to calculate the edge transform domain matrix $\mathbf{G}$ represents the discrete time wavelet smoothing function. Due to the inversion involved when obtaining $\mathbf{G}$, the wavelet function can not be freely dilated. Consequently, our method is not eligible for performing a multiscale/multi-resolution wavelet transform. The interest of a multi-resolution transform comes from the fact that the edges of interest would

show up always at the same positions. On the other hand, noise-induced spurious edges are random at each scale and thus tend not to propagate through all scales; hence, if a multiscale wavelet transform was available, an improved recovery could be applied.

- *Number of CR sensing receivers* (J): as the number of radios increases, the probability that every radio experiences a deep fade decreases. Hence, sensing reliability improves with the number of CRs. Due to this gain, sensing time, and thus averaging Q of individual radios can be reduced.

Note that there is a tradeoff between N and Q, i.e., a larger N improves the bandwidth resolution but at the same time increases the required averaging Q. In practice it is common to choose a fixed N to meet the desired resolution with a moderate complexity. Then, the number of spectral averages becomes the parameter used to meet the estimator performance goal.

5.2 Spectrum detection limitations

Our algorithms are based on some assumptions which are common in literature. Here we discuss them further:

1. We assumed that the noise is a white, additive and Gaussian wide-sense stationary process, with zero mean, and known variance. However, noise is an aggregation of various sources including not only thermal noise at the receiver and underlined circuits, but also interference due to nearby unintended emissions, weak signals from transmitters very far away, etc.

2. We assumed that noise variance is precisely known to the receiver, so that the threshold can be set accordingly. However, this is practically impossible as noise could vary over time due to temperature change, ambient interference, filtering, etc. The deviation from the assumed known value becomes particularly important when the signal strength is below the error of the noise variance. In that case, the detection threshold, which is set based on the known variance, is set too high and weak signals could never be detected.

3. In modeling the channels we assumed that they are independent and identically distributed. As a result, the diversity gains shown in the plots are maximized. However, channel coefficients are a result of superposition of three components (path loss, shadowing, and multipath) that do not necessarily need to be independent for all radios. While path loss for small to medium networks can be assumed equal for all radios, the other two effects could have quite different characteristics. For example, shadowing can exhibit high correlation if two radios are blocked by the same obstacle.

4. Our last assumption was made for equal distribution of noise and local interference across all radios. As a result, every CR could apply the same detection threshold and achieve equal p_{fa}. However, in all practical situations these two assumptions

do not hold. First, due to circuits variability or temperature difference, each radio has a different aggregate local noise.

In a nonstationary background, a fixed threshold NP detector cannot be used, because, as the background conditions vary, the resulting value of p_{fa} may be too high (i.e., decreasing the reuse of the unused spectrum) or the value of p_d may be too low (i.e., increasing the interference to the PU systems). This suggests to employ *adaptive threshold techniques* based on an estimate of the mean power level of the background noise. Several approaches may be reused from the noise estimation literature to solve the practical spectrum sensing problem: how can we set the threshold based on a real time estimation of the noise power, so that we can still guarantee the target probability of detection, or the probability of false alarm?

A simple but reasonable method is to treat the estimate of noise power as the true noise power and calculate the threshold used in energy detection accordingly. In [54], the practicality of performing real time noise estimation was justified with two examples:

1. Assume the spectrum regulators still want to reserve certain channels for special applications, and SUs are never allowed to access this channel. In addition, this special channel is rarely used and therefore can serve the purpose of noise estimation, e.g., in the United States, channel 37 (from 608 to 614 MHz) is reserved for radioastronomy and is used in very few occacions.

2. Detecting pilot signals which are distinct narrow band spectral features. After performing the PSD estimation on the received signal, the noise variance can be estimated from some frequency bin not corresponding to the pilot frequency.

Intuitively, this method may incur in an inherent loss of detection probability since the threshold is set by estimating the total noise power from only a finite number of observed noise samples. In [54] a method to determine the threshold from real time noise variance estimation is derived that can achieve the desired probability of detection of false alarm.

The previous approach relies on a noise variance estimated from a possibly vacant channel. In [55], the noise power is estimated from the full wideband signal with a method based on the shortest half sample, under the assumption of sparsity, i.e., the number of data points containing only noise is greater than the number of data points containing both signal and noise. The assumption is quite reasonable, since the current dynamic spectrum sharing research is motivated by the fact that many parts of the spectrum are under-utilized most of the times. This makes the concept of spectrum sharing to be attractive. Therefore, one can estimate the noise power without doing any explicit separation of the noise from the noisy signal. In [56] a wideband detector for single SUs or multiple collaborative SUs which does not require the noise variance is proposed based on the general likelihood ratio test (GLRT). They assume that among the subbands there is some minimum number of vacant subbands. Basically, the estimate of the noise variance in [56] is based on the average of the least energies of the subbands when sorted in an ascending order. This approach is intuitively justified, since it is more likely for the subbands with lower energies to be vacant rather than the higher energy ones.

Most of the fusion approaches in the literature have focused on the cases with conditionally i.i.d. observations. The correlated case where it is assumed that each SU knows the geographic locations of the other users and hence the correlation between the observations is studied in [57].

5.3 AIC implementation issues

Part of the CS research has focused on advanced devices for “analog-to-information” conversion (AIC) of high-bandwidth signals. The goal as introduced in this thesis is to alleviate the pressure on conventional ADC technology, which is currently limited to sampling rates on the order of 1 GHz.

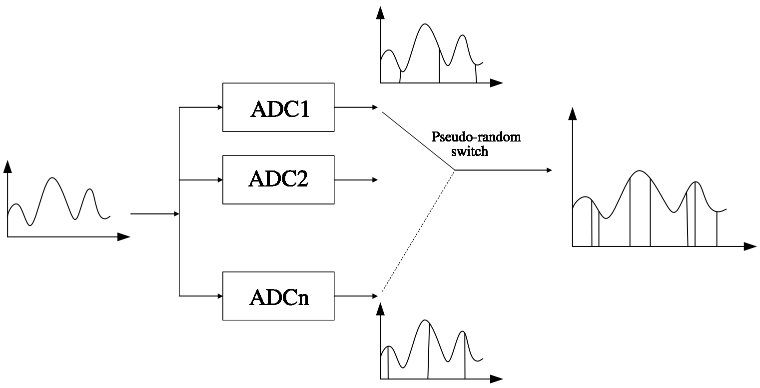

Figure 5.1: Random sampling scheme for AIC.

Probably because of the novelty of CS theory, there is not much literature about AIC hardware designs. We now provide a brief overview of existing methods in the actual literature:

- Random sampling
- Pseudo-random demodulation
- Random filtering.

In a nutshell, a low-rate sequence of measurements $y(m)$ can be acquired from a high bandwidth analog signal. Ideally, we would like to sample the signal at some multiple of the sparsity level, rather than at twice the bandwidth as demanded by the Nyquist sampling theorem.

The *random sampling* architecture digitizes the signal at randomly or pseudo-randomly sampled points. This architecture has been proposed for wideband signals that are sparse in a local Fourier representation in the sense that at each point in time they are well-approximated by a few local sinusoids of constant frequency (e.g., frequency hopping communication signals). Two implementations have been proposed in [50]. Fig. 5.1 depicts one of them. A bank of parallel low-rate ADCs that have equal shifts between their starting conversion points is used. This creates a shift in the samples that are produced from each of the parallel ADCs. The switching mechanism

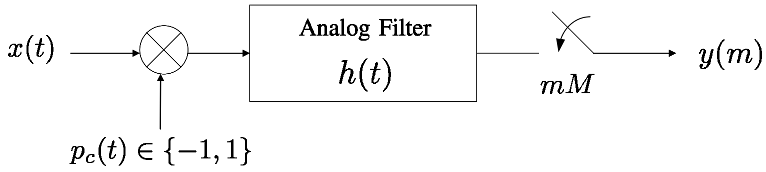

Figure 5.2: Pseudo-random demodulation scheme for AIC.

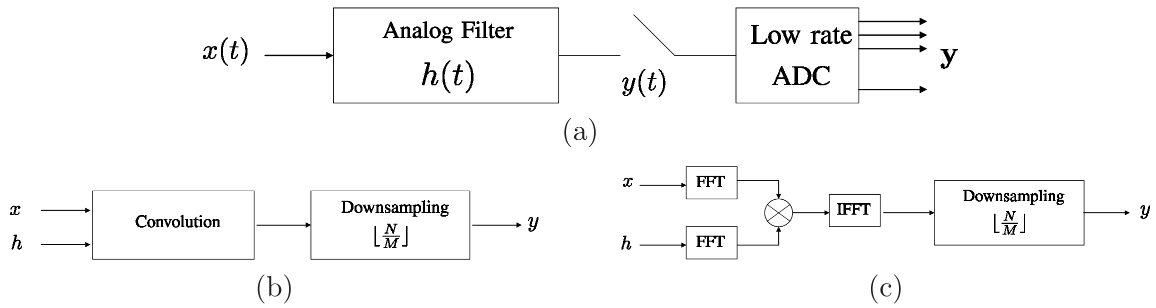

Figure 5.3: Random filtering scheme for AIC: (a) general scheme; (b) using convolution; (c) using FFT/IFFT. The FIR filter $\mathbf{h}$ has random taps, which must be known in order to recover the signal $\mathbf{x}$ from the compressed data $\mathbf{y}$.

is then controlled pseudo-randomly. The main implementation challenges come from the large chip area for the many ADCs and also the minimization of the jitter effect when controlling the switches.

The *pseudo-random demodulation* architecture [51] [52] is most notably applicable to those signals having a sparse representation in the time-frequency plane. Whereas it may not be possible to digitize an analog signal at a very high rate, it may be quite possible to change its polarity at a high rate. The basic idea is thus to multiply the signal by a pseudo-random sequence of $\pm$1s, integrate the product over time windows, and digitize the integral at the end of each time interval. The purpose of the demodulation is to spread the frequency content of the signal so that it is not destroyed by the integrator. This is a parallel architecture and one has several of these random multiplier-integrators pairs running in parallel using distinct sign sequences. In effect, this architecture correlates the signal with a bank of sequences of ± 1, one of the CS measurement processes known to be universal. Fig. 5.2 shows a schematic of the AIC based on pseudo-random demodulation. The most significant sources of non-idealities are: the clock jitter of the random number generator for $p_c(t)$, the linearity and intermodulation distortion of the mixer, and the quantization error of the back-end ADC.

The *random filtering* architecture (Fig. 5.3a) [53], in contrast to the previous two architectures is sufficiently generic to summarize many types of compressible signals sparse in the time, frequency, and wavelet domains, as well as piecewise smooth signals and Poisson processes (universal). It builds on the idea of random filters as a new paradigm for compressive signal acquisition. A random filter is an FIR filter whose taps are i.i.d. random variables. Seeking for consistency with the previous chapters, we are in particular interested in the case where the taps are drawn from the normal distribution. The wideband analog signal is hence captured by convolving it with a random-tap FIR filter $h(t)$ and then downsampling the filtered signal to obtain a compressed representation $\mathbf{y}$

$$\mathbf{y} = \mathcal{D}_{\downarrow}(\mathbf{h} \star \mathbf{x}) \tag{5.1}$$

where $\mathcal{D}_{\downarrow}$ downsamples by a factor of $\lfloor \frac{N}{M} \rfloor$. Note that, because this process is linear, the map from the signal $\mathbf{x}$ to the summary $\mathbf{y}$ can be viewed as $\mathbf{y} = \mathbf{\Phi x}$. Two implementations which we show in Fig. 5.3b and Fig. 5.3 have been proposed in [53].

5.4 Detection without estimation

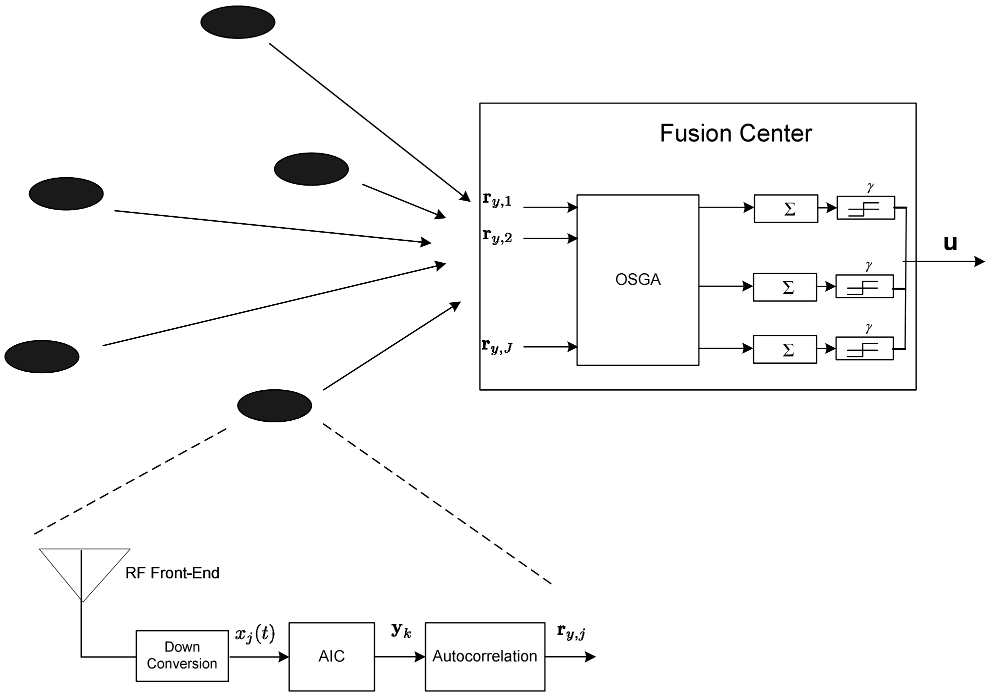

Figure 5.4: Block diagram of the architecture for OSGA.

In [38], a joint recovery algorithm has been proposed named as *One-Step Greedy Algorithm* (OSGA) intended to recover the joint common sparse support of a signal ensemble with fewer than $\mathcal{O}(K \log N)$ measurements per CR. Of course this approach does not recover the coefficients for each signal but it provides a sufficient statistic to perform detection at lower complexity, i.e. theorem 1 in [38] claims that with $M \geq 1$ measurements per signal, OSGA recovers the common sparse support with probability approaching 1 as $J \to \infty$. We next show from simulations that OSGA may be utilized as a recovery algorithm for the architecture shown in Figure 5.4 under the same assumption of joint sparsity concerning the case of multiple sparse signals that share common sparse components, but with different coefficients.

The measurements can be obtained following any of the acquisition schemes described in Chapter 3. Again, for simplicity we shall develop the equations for our proposed architecture (Section 3.3). We assume that an equal number of measurements is taken per CR and we write $\mathbf{\Theta}_j$ in terms of its columns

$$\mathbf{\Theta}_j = [\boldsymbol{\theta}_{j,1}, \boldsymbol{\theta}_{j,2}, \ldots, \boldsymbol{\theta}_{j,N}] \quad j = 1, \ldots, J, \tag{5.2}$$

where $\mathbf{\Theta}_j$ is given by

$$\mathbf{\Theta}_j = \mathbf{\Phi}_{II,j}\mathcal{F} \quad j = 1, \ldots, J, \tag{5.3}$$

with $\mathbf{\Phi}_{II,j}$ denoting the $2M \times 2N$ compressive sampling matrix as defined in (3.14) at the j-th CR and $\mathcal{F}$ standing for the $2N \times 2N$ discrete Fourier matrix. The measurements obtained under our proposed architecture in Section 3.3 follow

$$\mathbf{r}_{y,j} = \mathbf{\Phi}_{II,j}\mathcal{F}\mathbf{S}_{x,j} \quad j = 1, 2, \ldots, J, \tag{5.4}$$

with $\mathbf{S}_{x,j}$ denoting the PSD of the sought signal $x_j(t)$ at the j-th CR. After gathering all of the measurements the following statistic is computed

$$\xi_n = \frac{1}{J\chi}\sum_{j=1}^{J}\langle \mathbf{r}_{y,j}^T, \boldsymbol{\theta}_{j,n}\rangle^2 \quad n = 1, 2, \ldots, 2N, \tag{5.5}$$

where χ denotes the mean of the test statistic ξ_n in the absence of the PU signal.

In [39] the mean and the variance of ξ_n have been found assuming Gaussian measurement matrices, i.e. $[\mathbf{\Phi}_{II,j}]_{m,n} \sim \mathcal{N}(0, 1)$, Gaussian signal entries, i.e. $\mathbf{S}_{x,j} \sim \mathcal{N}(0, \sigma^2)$ and a common sparse support denoted by Ω with $|\Omega| = K$. They show that under the above assumptions the mean and the variance of ξ_n are given by

$$E(\xi_n) = \begin{cases} m_b, & \text{if } n \neq \Omega \\ m_g, & \text{if } n \in \Omega \end{cases} \tag{5.6}$$

and

$$Var(\xi_n) = \begin{cases} \sigma_b^2, & \text{if } n \neq \Omega \\ \sigma_g^2, & \text{if } n \in \Omega \end{cases} \tag{5.7}$$

where

$$m_b = MK\sigma^2, \tag{5.8}$$

$$m_g = M(M + K + 1)\sigma^2, \tag{5.9}$$

$$\sigma_b^2 = \frac{2MK\sigma^4}{J}(MK + 2K + 3M + 6), \tag{5.10}$$

$$\sigma_g^2 = \frac{M\sigma^4}{J}(34MK + 6K^2 + 28M^2 + 92M + 48K + 90 + 2M^3 + 2MK^2 + 4M^2K). \tag{5.11}$$

We note that the background level depends on the unknown sparsity of the signal and also on the parameters of the unknown distribution of the signal. Hence, we propose to normalize the algorithm by including the normalization factor χ, consisting of an estimate of the background level, which allows us to define a predetermined detection threshold independent of the unknown parameters of the received signal. The parameter χ may be estimated with any of the methods discussed in Section 5.2.

We propose two methods for decentralized computation in case a FC is not available

- *Method 1*: Each CR obtains its own $\mathbf{r}_{y,j}$ and then computes

$$\langle \mathbf{r}_{y,j}, \boldsymbol{\theta}_{j,n} \rangle \quad n = 1, 2 \ldots, 2N. \tag{5.12}$$

  Afterwards, average consensus is applied to obtain an estimate of ξ_n. This method requires every CR to share $2N$ samples.

- *Method 2*: Every CR broadcasts its measurement vector $\mathbf{r}_{y,j}$ ($2M$ measurements). After gathering all measurements from all neighboring nodes, each CR may compute an estimate of ξ_n.

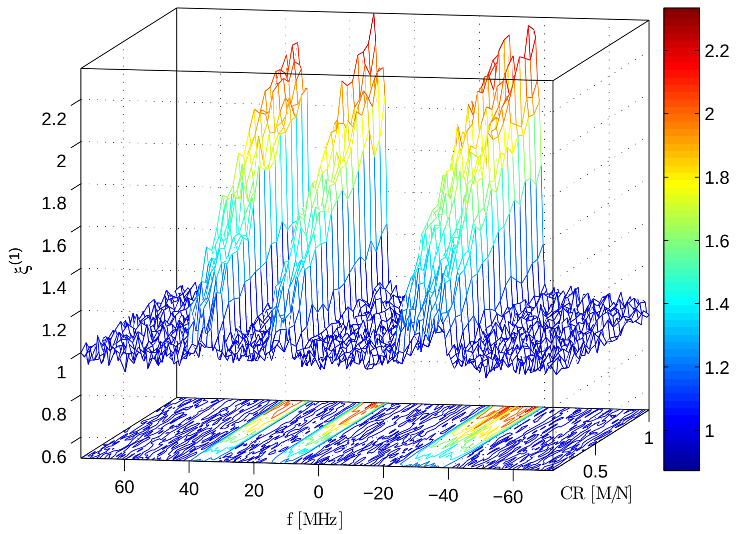

Figure 5.5: $\boldsymbol{\xi}^{(1)}$ for different compression ratios when $J = 1000$.

Let's denote the $2N \times 1$ test statistic vector as

$$\boldsymbol{\xi} = [\xi_1 \ \xi_2 \ \ldots \ \xi_{2N}]. \tag{5.13}$$

Next we plot $\boldsymbol{\xi}$ for the test signal introduced in Section 3.5 with a new occupancy ratio of 40%, i.e., 4 out of 10 subbands are occupied by licensed transmission signals and the remaining 6 channels are unoccupied. The received SNR of the 4 active channels is uniformly varied ranging from -5 to 5 dB and N is set to 128. We shall estimate $\boldsymbol{\xi}$ for both schemes described in Chapter 3. Hence in the simulations the test statistic under the acquisition scheme of Section 3.2 is denoted by $\boldsymbol{\xi}^{(1)}$ and the one under the scheme proposed in Section 3.3 is denoted by $\boldsymbol{\xi}^{(2)}$.

Simulation scenario 1: Figs. 5.5 and 5.6 show $\boldsymbol{\xi}^{(i)}$, $i = 1, 2$, when the number of CRs is fixed to $J = 1000$ and the compression rate is set to vary from 1% to 100%. As expected,$\boldsymbol{\xi}^{(i)}$, $i = 1, 2$ follow the occupancy of the subbands.

Simulation scenario 2: Figs. 5.7 and 5.8 show again $\boldsymbol{\xi}^{(i)}$, $i = 1, 2,$, when the compression rate is 0.05 and the number of CRs is set to vary from 1 to 5000. Note that the proposed algorithm of Section 4.1.2.1 would completely fail to recover the sparse signal for such a low compression rate. Fig. 5.9 shows a snapshot of Figs. 5.5

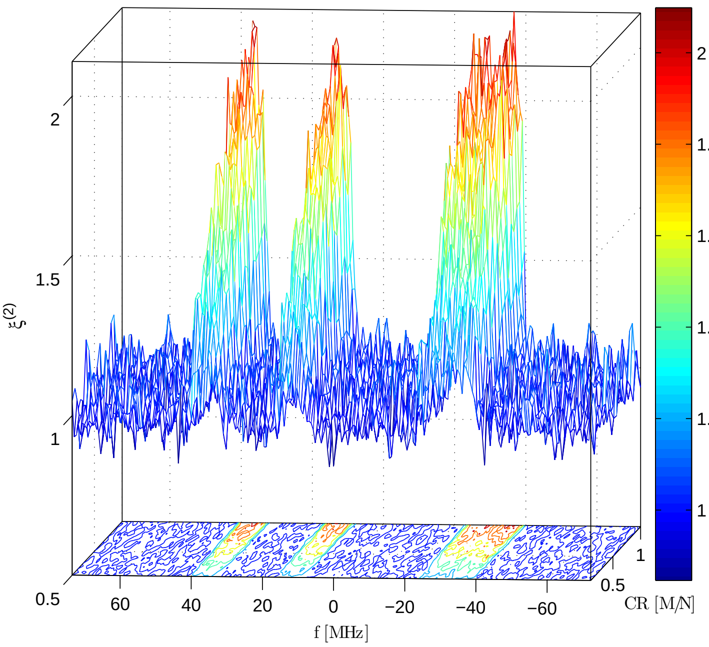

Figure 5.6: $\boldsymbol{\xi}^{(2)}$ for different compression ratios when $J = 1000$.

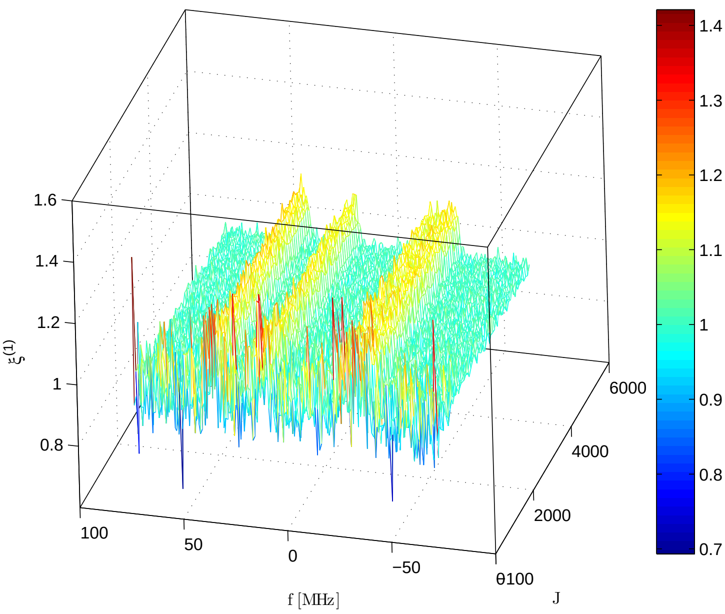

Figure 5.7: $\boldsymbol{\xi}^{(1)}$ for different J when $CR = 0.05$.

and 5.6 for a fixed compression rate of 0.05. The top plot shows one realization of the original PSD of the received wide-band signal where the signal to noise ratios of the 4

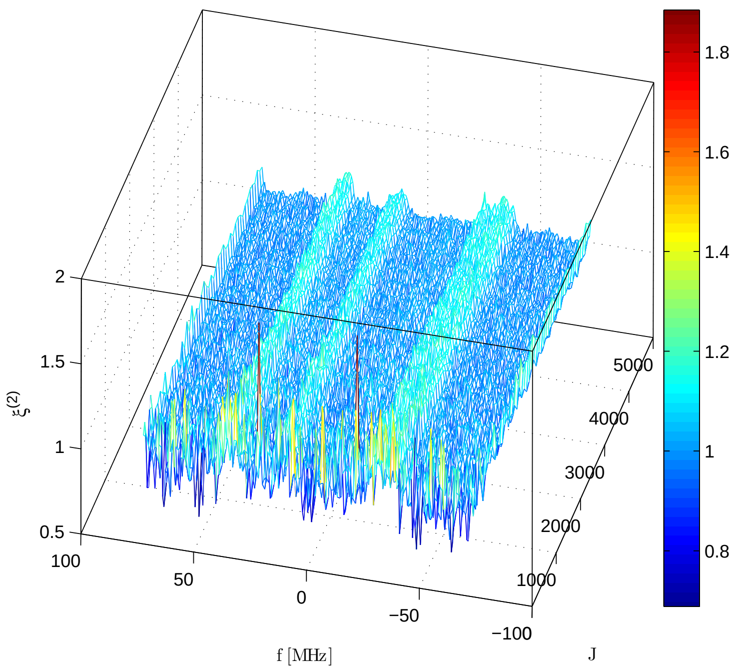

Figure 5.8: $\boldsymbol{\xi}^{(2)}$ for different J when $CR = 0.05$.

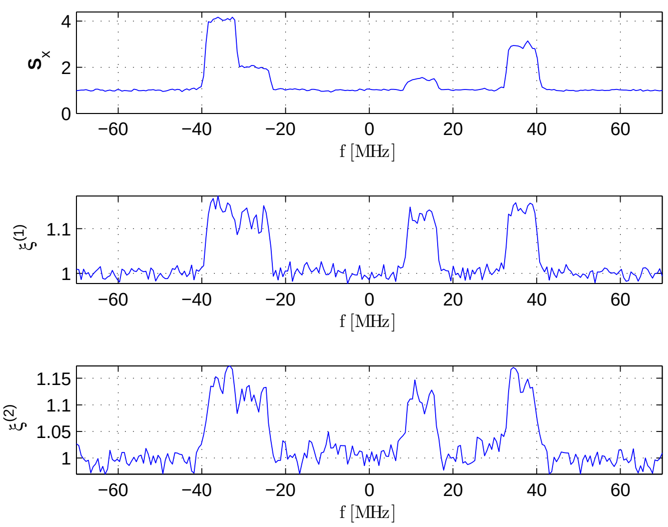

Figure 5.9: (a) Nyquist rate PSD with SNR equal to 5, 0, -3 and 3 dB; (b) $\boldsymbol{\xi}^{(1)}$; (c) $\boldsymbol{\xi}^{(2)}$

active channels are 5 dB, 0 dB, -3 dB and 3 dB, respectively.

Simulation scenario 3: In Fig. 5.10 we evaluate the probability of detection P_D based on the statistic $\boldsymbol{\xi}^{(2)}$. The decision of the presence of a licensed transmission signal in a certain subband is again made by the energy detector of (3.26). The normalization

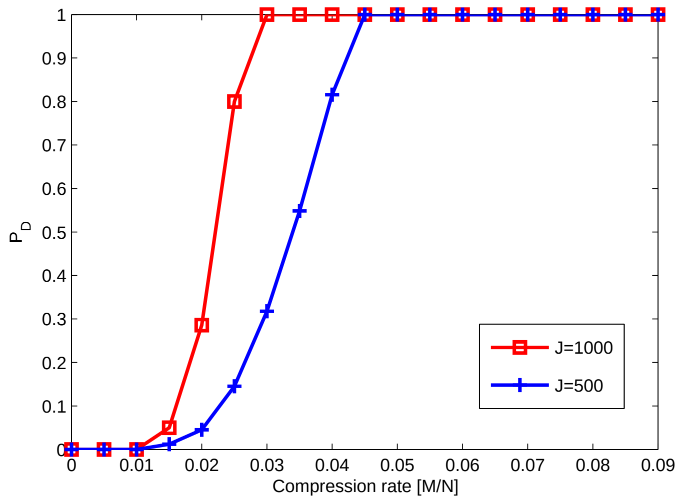

Figure 5.10: Detection performance

factor χ is estimated from a vacant subchannel following

$$\chi = \frac{1}{K} \sum_{k=(p-1)K}^{pK-1} \xi^{(2)}(k) \tag{5.14}$$

where p refers to the vacant channel and $K = 14$. Note again that the probability of detection reaches 1 for very small compression rates as long as the number of sensors is large enough.

All figures show the averaging behavior of the statistic $\boldsymbol{\xi}^{(2)}$, i.e., the SNR of the active channels is averaged out providing the algorithm with robustness against deep fades.

5.5 Conclusions

We have discussed several practical limitations in the context of compressive sensing in CRs and we have described the main architectures available in literature to implement the AIC. While potentially practically useful, the discussed algorithm OSGA requires J to be large. Our numerical experiments predict a good performance when M is small, as long as J is sufficiently large. However, in the case of fewer signals (small J), OSGA performs poorly.

6 Conclusions and further work

In this chapter we summarize the work done in the thesis, draw the final conclusions, and suggest directions for further research.

6.1 Conclusions

We presented fundamental schemes for wideband spectrum sensing in CRs based on compressive sampling. We addressed problems with current state-of-art approaches that require a wideband ADC, by proposing a scheme that works at sub-Nyquist rates. Several architectures and algorithms were provided depending on whether or not there was coordination between different CRs in the CR network.

In Chapter 3 we presented a spectrum sensing scheme based on CS for applications on wideband CR systems. The main advantage of this scheme is that it operates well below Nyquist rate. We compared its performance with the energy detector and it gives similar performance. Note that each CR is designed not expecting cooperation from other users in the detection process. Actually, in practice, cooperation between the CR users cannot be guaranteed in general, since a user can cooperate with others only when there are also other users in its vicinity monitoring the spectrum. A more feasible and reliable system where the individual secondary users make independent decisions about the presence of the primary users and communicate their decisions to a fusion center was considered in Section 4.1.1 (independent CS). The fusion center makes the final decision about the occupancy of the band by fusing the decisions made by all distributed radios in that area. In practice, the fusion center could be some centralized controller that manages the channel assignment and scheduling for the secondary users. In the same section we also proposed a solution for the topology where a centralized fusion center is not available, and where the secondary users exchange their decisions and each secondary user performs its own fusion of all the decisions.

The algorithms proposed in Sections 4.1.2 (joint CS) and 5.4 (OSGA) rely on the existence of a distributed network of CR sensing receivers. While the joint CS architecture does not need a large number of secondary users to provide good detection results, OSGA requires a very large number of low complexity sensing devices, subsequently, it suits applications where a dense sensor network is available. In a dense sensor network, a potentially large number of distributed sensor nodes can be programmed to perform data acquisition tasks as well as to network themselves to communicate their results to a central collection point. The main advantage of our architectures rely on the capability of giving the same results as existing systems from presumably less samples.

Finally, in Chapter 5 we provided an extensive discussion about the major practical implementation issues.

6.2 Suggestions for further Work

- **Sequential compressive sensing**
  Existing analytical results on CS provide guidelines on how many measurements are needed to ensure exact recovery with high probability, but these are often seen to be pessimistic and rely on a priori knowledge about the sparsity of the unknown signal. A more suitable scenario would then be to get observations in sequence, and perform computations in between observations to decide whether enough samples have been obtained. Exact recovery would be in that case, possible from the smallest possible number of observations, and without any a priori knowledge about how sparse the underlying signal is.

- **Direct detection from compressive measurements**
  Our research has focused on the reconstruction of the PSD (relying on the assumption of sparsity). However, seeking to reduce the complexity of the algorithms involved, we must not forget that the fundamental task is not estimating the PSD but detecting the presence of the primary users, subsequently, the full reconstruction of the signal should not be required. In Section 5.4, we described a first approach meeting this goal, without going into the intermediate stage of estimating the PSD.

- **CS recovery algorithm**
  When the wavelet transform is involved, a multi-resolution solution may be available by dilating the wavelet basis function. The interest in a multi-resolution transform comes from the fact that the edges of interest would show up always at the same positions for different scalings, however, noise-induced spurious edges are random at each scale and thus tend not to propagate through all scales; hence, if a multiscale wavelet transform was available, an improved recovery could be applied.

- **Decentralized computation**
  When performing joint recovery in a distributed wireless network under the common sparse support assumption when no CSI is available, we relied on a fusion center. It may be interesting trying to completely distribute the computation of a detection consensus from compressive measurements.

Nonparametric and parametric spectrum estimation techniques review

A

Here is a short review of the parametric and the non parametric methods following the presentation in [18].

A.1 Nonparametric estimation

One way of estimating the PSD of a process is to simply find the discrete time Fourier transform of the samples of the process (usually done with an FFT) and take the magnitude squared of the result. This estimate is called the *periodogram.* The periodogram of a length N signal $x(n)$ is

$$\hat{S}_x(f) = \frac{|X(f)|^2}{N} \tag{A.1}$$

where

$$X(f) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn}. \tag{A.2}$$

Since $\hat{S}_x(f)$ is a function of random variables, it is necessary to consider convergence in a statistical sense. Therefore we are interested in whether or not

$$\lim_{N\to\infty} E\{[\hat{S}_x(f) - S_x(f)]^2\} = 0. \tag{A.3}$$

In order for the periodogram to be mean-square convergent, it is necessary that it is asymptotically unbiased

$$\lim_{N\to\infty} E\{[\hat{S}_x(f)]\} = S_x(f) \tag{A.4}$$

and have a variance that goes to zero as the data record length N goes to infinity,

$$\lim_{N\to\infty} Var\{\hat{S}_x(f)\} = 0. \tag{A.5}$$

In other words, $\hat{S}_x(f)$ must be a consistent estimate of the power spectrum.

The periodogram is a biased estimator of the PSD. Its expected value can be shown to be

$$E[\hat{S}_x(f)] = \sum_{m=-(N-1)}^{N-1} (1 - \frac{|m|}{N})r_x(m)e^{-j2\pi fm}. \tag{A.6}$$

This suggests that the estimates produced by the periodogram corresponds to a *leaky* PSD rather than the true PSD. Note that $1 - \frac{|m|}{N}$ essentially yields a triangular Bartlett window (which is apparent from the fact that the convolution of two rectangular pulses is a triangular pulse). This results in leakage, distortion increasing the sidelobes, and smoothing (decresased resolution). However, it is asymptotically unbiased, which is evident from the earlier observation that as the data record length tends to infinity, the frequency response of the rectangular window more closely approximates the Dirac delta function (also true for a Bartlett window). However, in some cases the periodogram is a poor estimator of the PSD even when the data record is long. This is due to the variance of the periodogram, i.e., the periodogram is not a consistent estimator of the PSD: increasing the number of samples increases frequency resolution but does not enhance the variance of the estimate. There are some nonparametric methods that are able to reduce the variance in the spectral estimate by decreasing the frequency resolution. Then the quality of the power spectrum estimate increases when sample size is increased.

The *modified periodogram* windows the time-domain signal prior to computing the FFT in order to smooth the edges of the signal

$$\hat{S}_x(f) = \sum_{m=-(N-1)}^{N-1} x(m)w(m)e^{-j2\pi fm}. \tag{A.7}$$

Common used windows are Hann and Hamming. This has the effect of reducing the spectral leakage. However it also results in a reduction of resolution.

In the *Bartlett method* the N-point sequence is subdivided into K non-overlapping segments. The periodogram is computed for each segment and the periodograms are averaged.

$$\hat{S}_x(f) = \frac{1}{K}\sum_{i=0}^{K-1}\frac{1}{M}|\sum_{n=0}^{M-1} x(n+iM)e^{-j2\pi fn}|^2. \tag{A.8}$$

Variance of the estimate is then reduced by factor of K.

The *Welch method* consists of dividing the time signal into overlapping segments, computing a modified periodogram of each segment, and then averaging the PSD estimates

$$\hat{S}_x(f) = \frac{1}{L}\sum_{i=0}^{L-1}\frac{1}{MU}|\sum_{n=0}^{M-1} x(n+iD)w(n)e^{-j2\pi fn}|, \tag{A.9}$$

where

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n). \tag{A.10}$$

Each periodogram is calculated starting from sample iD. Thus, the Welch method is the same than the Bartlett method if $D = M$ and the window function $w(n)$ is constant. The variance of Welch's estimator is difficult to compute because it depends on both the window used and the amount of overlap between segments. Basically, the variance is inversely proportional to the number of segments whose periodograms are being averaged. In summary there is a trade-off between variance reduction and resolution.

The methods of Bartlett and Welch are designed to reduce the variance of the periodogram by averaging periodograms and modified periodograms. Another method for decreasing the statistical variability of the periodogram is periodogram smoothing, often referred to as the Blackman-Tukey method. In the *Blackman and Tukey method* the sample autocorrelation sequence is windowed first and then Fourier transformed in order to decrease the contribution of the unreliable autocorrelation estimates

$$\hat{S}_x(f) = \sum_{m=-(M-1)}^{M-1} \sum_{n=-\infty}^{\infty} x^*(n)x(n+m)w(m)e^{-j2\pi fm}. \tag{A.11}$$

Here the window function $w(n)$ has length $2M-1$ and is zero for $|m| \geq M$.

The periodogram can be interpreted as filtering a length N signal, $x(n)$, through a filter bank of N FIR bandpass filters. The magnitude response of each one of these bandpass filters resembles that of the rectangular window. The periodogram can thus be viewed as a computation of the power of each filtered signal, i.e., the output of each bandpass filter, that uses just one sample of each filtered signal and assumes that the PSD of $x(n)$ is constant over the bandwidth of each bandpass filter. As the length of the signal increases, the bandwidth of each bandpass filter decreases, making it a more selective filter, and improving the approximation of constant PSD over the bandwidth of the filter. This provides another interpretation of why the PSD estimate of the periodogram improves as the length of the signal increases. However, there are two factors apparent from this standpoint that compromise the accuracy of the periodogram estimate. First, the rectangular window yields a poor bandpass filter. Second, the computation of the power at the output of each bandpass filter relies on a single sample of the output signal, producing a very crude approximation.

Welch's method can be given a similar interpretation in terms of a filter band. In Welch's implementation, several samples are used to compute the output power, resulting in reduced variance of the estimate. On the other hand, the bandwidth of each bandpass filter is larger than that corresponding to the periodogram method, which results in a loss of resolution. The filter bank model thus provides a new interpretation of the compromise between variance and resolution.

Multitaper method (MTM) [19] and [20] builds on these results to provide an improved PSD estimate. In MTM the power spectrum estimate for a frequency slot $[f_i - \Delta, f_i + \Delta]$ is calculated by averaging over output of several filters or tapers. This reduces the variance of the estimate. The tapers are orthogonal to each other and are centered on the frequency f_i. The filters are called Slepian sequences $\{w_t^{(k)}\}_{t=1}^{N}$. The data $x(t)$ is expanded using the Slepian sequencies and an eigenspectrum Y_k is calculated according to

$$Y_k(f) = \sum_{t=1}^{N} w_t^{(k)} x(t) e^{-j2\pi ft}. \tag{A.12}$$

The power spectrum estimate consists of a weighted sum of the eigenspectra

$$\hat{S}_x(f) = \frac{\sum_{k=0}^{K-1} \lambda_k(f)|Y_k(f)|^2}{\sum_{k=0}^{K-1} \lambda_k(f)}, \tag{A.13}$$

where λ_k is the eigenvalue corresponding to eigenvector with elements of the k-th taper $\{w_t^{(k)}\}_{t=1}^{N}$. The more concentrated the taper is to the frequency slot $[f_i - \Delta, f_i + \Delta]$ the higher the value of λ_k.

In [21], they show that for wideband signals, the multitaper spectral estimation procedure is "nearly optimal" in the sense that it almost achieves the Cramer-Rao bound for a nonparametric spectral estimator.

However, nonparametric methods are based on non-realistic assumptions. There is an inherent assumption that the autocorrelation estimate is zero for samples $> N$. Also, there is an inherent assumption that the data is periodic with period N. The assumptions limit frequency resolution and quality of the estimate. The power spectrum can be estimated without these assumptions using parametric methods. They result in better frequency resolution and better quality of estimate with finite sample size.

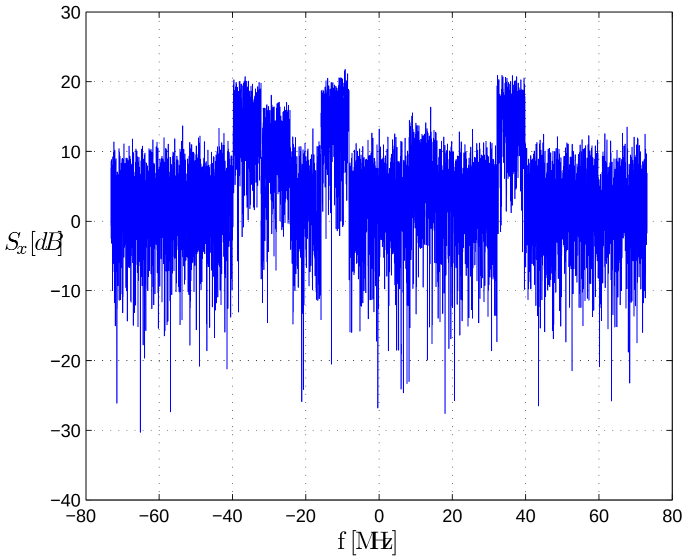

Figure A.1: Periodogram wideband estimation method.

A.2 Parametric estimation

A different limitation of the nonparametric methods is that they are not designed to incorporate information that may be available about the process into the estimation procedure. In some applications this may be an important limitation, particularly when some knowledge is available about how the data samples are generated. Parametric methods can yield higher resolutions than nonparametric methods in cases when the signal length is short. These methods use a different approach to spectral estimation; instead of trying to estimate the PSD directly from the data, they model the data as the output of a linear system driven by white noise, and then attempt to estimate the parameters of that linear system

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + \sum_{k=0}^{q} b_k w(n-k). \tag{A.14}$$

With a parametric approach, the first step is to select an appropriate model for the

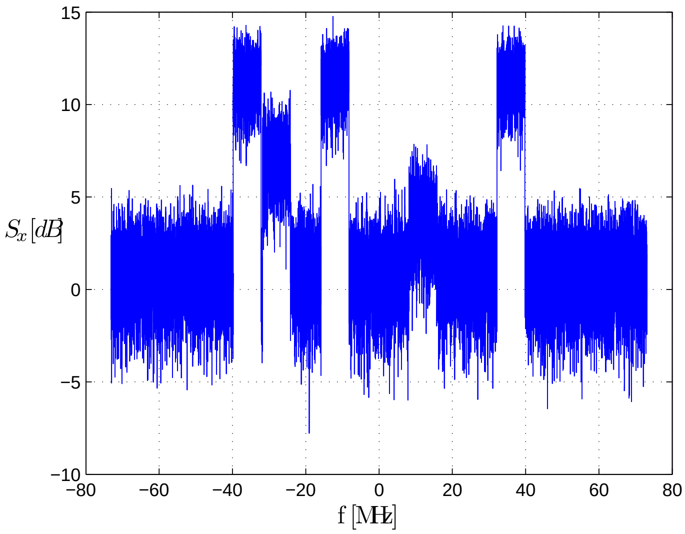

Figure A.2: Welch wideband estimation method.

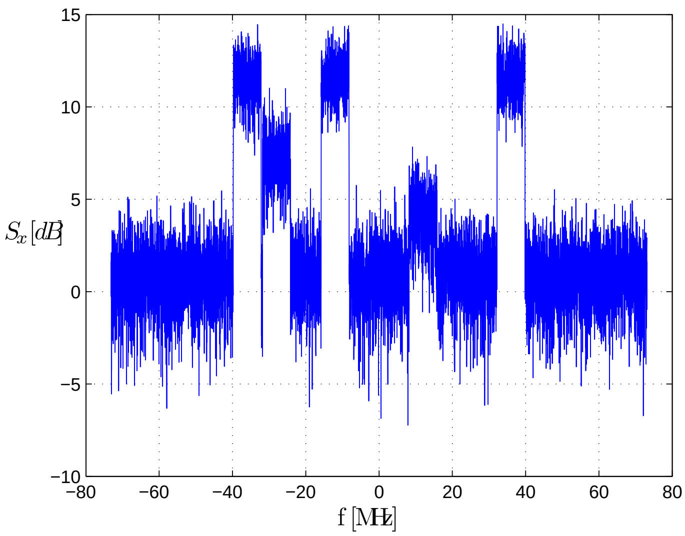

Figure A.3: Multitaper wideband estimation method

process. This selection may be based on a priori knowledge about how the process is generated or, perhaps on experimental results indicating that a particular model "works well". Models that are commonly used include *autoregressive* (AR), *moving average* (MA), and *autoregressive moving average* (ARMA).

In ARMA the data sequence is the output of a linear system characterized by a system function

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}}. \tag{A.15}$$

In AR the system function is

$$H(z) = \frac{1}{A(z)}. \tag{A.16}$$

The AR model is suitable for representing spectra with narrow peaks. It is also the most commonly used parametric method since it results in simple linear equations for the a_k parameters. The MA is characterized by

$$H(z) = B(z). \tag{A.17}$$

The power density spectrum is calculated from the system function using

$$\hat{S}_x(f) = \sigma_w^2 |H(f)|^2. \tag{A.18}$$

Although it is possible to significantly improve the resolution of the spectrum estimate with a parametric method, it is important to realize that, unless the model that is used is appropriate for the process that is being analyzed, inaccurate or misleading estimates may be obtained.

A.3 Simulation results

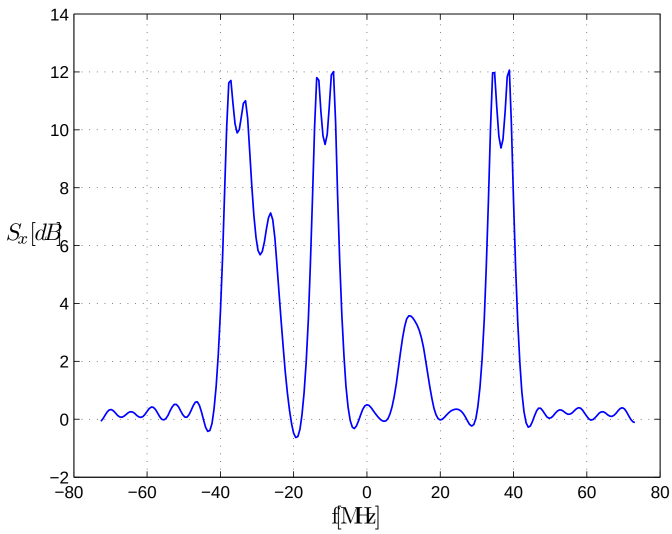

Figure A.4: AR wideband estimation method

The PSD is estimated for the test case introduced in Section 3.5. We consider, at baseband, a wide frequency band of interest ranging from -40 to 40 MHz, containing $P = 10$ non-overlapping subbands of equal bandwidth of 8 MHz. 5 subbands are occupied by signals using OFDM modulation according to the DVB-T standard. Each 8 MHz OFDM symbol has 8192 frequency tones and a cyclic prefix length of 1024. The number of OFDM symbols used for spectrum sensing is 2. The over-sampling factor is 16. The signal whose PSD is to be estimated is corrupted by additive white Gaussian noise (AWGN) with a variance of $\sigma_n^2 = 1$. The received signal to noise ratios (SNR) of the 5 occupied channels are 10dB, 5dB, 10dB, 0dB, and 10dB, respectively. Spectrum estimation was performed with four different methods: periodogram, Welch method, AR method, and multitaper method. The parameters for the methods are

- Periodogram (Fig. A.1): the FFT length considered is N=8192.
- Welch method (Fig. A.2): the number of segments is set to $K = 8$,with an overlapping of 50% and a Hamming windowing.
- Multitaper method (Fig. A.3): the number of discrete prolate spheroidal sequences (Slepian sequences) is set to 7.
- AR model (calculated with Yule-Walker method) (Fig. A.4): the order of the autoregressive model is set to $p = 25$.

Bibliography

[1] Shared Spectrum Company,"Spectrum occupancy measurements", from www.sharedspectrum.com

[2] Federal Communications Commission - First Report, and Order and Further Notice of Proposed Rulemaking, "Unlicensed operation in the TV broadcast bands," *FCC 06-156*, Oct. 2006.

[3] FCC, "Spectrum policy task force report," *ET. Docket No. 02-155*, Nov. 2002.

[4] J. Mitola,"Cogntive radio for flexible mobile multimedia communications," *Proc. IEEE Int. Workshop on Mobile Multimedia Communication*, pp. 3-10, 1999.

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

[6] A. Goldsmith, S.A. Jafar, I. Maric, S. Srinivasa, "Breaking the spectrum gridlock with cognitive radios: an information theoretic perspective," *Proceedings of the IEEE*, invited, to appear 2008.

[7] D. Cabric, I.D. O'Donnell, M.S. Chen, and R.W. Brodersen, "Spectrum sharing radios", *IEEE Circuits and Systems Magazine*, pp. 30-45, 2006.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall, 1998.

[9] A. Sahai and D. Cabric, "A tutorial on spectrum sensing: fundamental limits and practical challenges," *Proc. IEEE Symposium on New Frontiers in Dynamic Spectum Access Networks* (DySpan), Baltimore, MD, Nov. 2005.

[10] B. Farhang-Boroujeny, "Filter bank spectrum sensing for cognitive radios," in *IEEE Transactions on Signal Processing*, vol. 56, no. 5, May 2008.

[11] M. Davenport, M. Wakin, and R. Baraniuk, "Detection and estimation with compressive measurements." Rice ECE Department Technical Report TREE 0610, November 2006.

[12] J. Haupt and R. Nowak, "Compressive sampling for signal detection," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP), Honolulu, Hawaii, April 2007.

[13] D. Cabric, A. Tkachenko, and R.W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," *Proc. ACM Int. Workshop on Technology and Policy for Accessing Spectrum* (TAPAS), Boston, MA, Aug. 2006.

[14] J. Baek, H. Oh, and S. Hwang, "Improved reliability of spectrunm sensing using energy detector in cognitive radio systems," *10th International Conference on Advanced Communication Technology* (ICACT), Feb. 2008.

[15] H. Chen, W. Gao, and D.G. Daut, "Spectrum sensing for OFDM systems employing pilot tones and application to DVB-T OFDM," *IEEE International Conf. on Communications* (ICC), May 2008.

[16] C. Cordeiro, M. Ghosh, D. Cavalcanti, and K. Challapali, "Spectrum sensing for dyanamic spectrum access of TV bands," *2nd Int. Conf. on Cognitive Radio Oriented Wireless Networks and Communications* (CrownCom), Aug. 2007.

[17] C. R. C. M. da Silva, B. Choi, and K. Kim, "Distributed spectrum sensing for cognitive radio systems," *Information Theory and Applications Workshop*, Oct. 2007.

[18] M.H. Hayes, *Statistical digital signal processing and modelling*, John Wiley and Sons, New York, April 1996.

[19] T. Walden, "Multitaper estimation of the innovation variance of a stationary time serie," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 181-187, 1995.

[20] S. Haykin, "Cognitive radio: brain-empowered wireless communication," *IEEE J. Selected Areas of Communications*, vol. 23, no. 2, pp. 201-220, 2005.

[21] P. Stoica, and T. Sundin, "On nonparametric spectral estimation," *Circuits, Syst., Signal Process.,*, vol. 16, pp. 169-181, 1999.

[22] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[23] E. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.

[24] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, 23(3), pp. 969-985, June 2007.

[25] [7] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 12071223, 2006.

[26] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," *Proc. of the International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), pp. IV/1357-IV/1360, April 2007.

[27] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Information Theory*, vol. 52, no. 1, pp. 618, 2006.

[28] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signal," *IEEE Trans. Info. Theory*, vol. 51, no. 3, pp. 10301051, 2006.

[29] S. S. Chen, D. L. Donoho, and M. A. Saunders. "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 43, no. 1, pp. 129159, 2001.

[30] C. La and M. Do, “Signal reconstruction using sparse tree representations,” *SPIE Wavelets XI*, vol. 5914, pp. 59140W.1-59140W.11, Sept. 2005.

[31] P. J. Huber, “Projection pursuit,” *The Annals of Statistics*, vol. 13, pp. 435475, 1985.

[32] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad., “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” *Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, 1993.

[33] I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm,” *IEEE Trans. Signal Process.*, vol. 45, pp. 600 616, Mar. 1997.

[34] R. Chartrand and W. Yin, “Iteratively reweighted algorithms for compressive sensing,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, (ICASSP), Las Vegas, NV, pp. 3869 3872, April 2008.

[35] A. E. Hoerl and R. W. Kennard, “Ridge regression: biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, pp. 5567, 1970.

[36] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267 288, 1996.

[37] T. Tao, “Open question: deterministic UUP matrices,” July 2007. http://terrytao.wordpress.com/2007/07/02/open-questiondeterministic-uup-matrices/.

[38] M.F. Duarte, S.. Sarvotham, D. Baron, M.B. Wakin, R.G. Baraniuk “Distributed compressed sensing of jointly sparse signals,” *Thirty-Ninth Asilomar Conference*, pp 1537-1541, Oct. 2005.

[39] S. Sarvotham, M.B. Wakin, D. Baron, M.F. Duarte, and R.G. Baraniuk, “Analysis of the DCS one-stage greedy algoritm for common sparse supports,” Tech. Rep., Rice University ECE Department, Oct. 2005, available at http://www.ece.rice.edu/ shri/docs/TR0503.pdf.

[40] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solution to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Processing*, vol.51, pp. 2477-2488, July 2005.

[41] Z. Tian, E. Blasch, W. Li, G. Chen, and X. Li, “Performance evaluation of distributed compressed wideband sensing for cogntiive radio networks,” *11th Int. Conf. on Information Fusion*, Cologne, July 2008.

[42] W. Zhang, R.K. Mallik, K.B. Letaief “Cooperative spectrum sensing optimization in cognitive radio networks,” *IEEE Int. Conf. on Communications* (ICC), pp 3411-3415, May 2008.

[43] P.K. Varshney, “Distributed detection and data fusion”, Springer-Verlag, New York, 1997.

[44] D. Cabric, S.M. Mishra, and r. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," *Proc. 38th Asilomar Conference on signals, Systems and Computers*, Pacific Grove, CA, Nov. 2004.

[45] S. Haykin, "Cognitive radio: Brain-empowered wireless communicaciones," *IEEE J. Select. Areas Commun.,*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[46] F.F. Digham, M. Alouini, and M.K. Simon, "On the energy detection of unknown signals over fading channels" *IEEE Int. Conf. on Communications* (ICC), May 2003.

[47] Z. Quan, S. Cui, A.H. Sayed, and H.V. Poor,"Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks," *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), Las Vegas, NV, April 2008.

[48] A.D.G. Dimakis, A.D. Sarwate, and M.J. Wainwright, "Geographic gossip: efficient averaging for sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, March 2008.

[49] L. Xiao, S. Boyd, and S. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33-46, 2007.

[50] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss, "Random sampling for analog-to-information conversion of wideband signals," *IEEE Dallas Circuits and Systems Workshop*, pp. 119-122, Oct. 2006.

[51] S. Kirolos, T. Ragheb, J. Laska, M.F. Duarte, Y. Masssoud, and R. G. Baraniuk, "Practical issues in implementing analog-to-information converters," *IEEE International Workshop on System on Chip for Real Time Applications*, pp. 141-146, Dec. 2006.

[52] M. Mishali, Y.C. Eldar, J.A. Tropp, "Efficient sampling of sparse wideband analog signals," Tech. Rep., CCIT Report # 705, Oct. 2008.

[53] J.A. Tropp, M.B. Wakin, M.F. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proc. Int. Conf. Acoustics, Speech and Signal Processing* (ICASSP), May 2006.

[54] Z. Ye, G. Memik, J. Grosspietsch, "Energy detection using estimated noise variance for spectrum sensing in cognitive radio networks," *Wireless Communication and Networking Conference* (WCNC), April 2008.

[55] D. Makovoz, "Noise variance estimation in signal processing," *IEEE Int. Symposium on Signal Processing and Information Technology*, Aug. 2006.

[56] A. Taherpour, S. Gazor, and M. Nasiri-Kenari, "Wideband spectrum sensing in unknown white Gaussian noise," *Special issue on Cognitive Spectrum Access in IET Communications*, Vol. 2, No. 6, pp.763-771, 2008.

[57] J. Unnikrishnan, and V.V. Veeravalli, "Cooperative spectrum sensing and detection for cognitive radio," *IEEE Global Communications Conference* (GLOBECOM), Nov. 2007.

[58] A.C. Gurbuz, J.H. McClellan, and V. Cevher, "A compressive beamforming method," *Proc. Int. Conf. Acoustics, Speech and Signal Processing* (ICASSP), 2008.

[59] R. Hleiss, P. Duhamel, M. Charbit, "Oversampled OFDM systems," *13th Inter. Conf. on Digital Signal Processing* (DSP), 1997.

[60] R. Baraniuk, "Compressive Sensing," Lecture notes in *IEEE Signal Processing Magazine* Volume 24, July 2007.