Compressive Detection and Estimation with Applications to Cognitive Radio and Radar

Linda Bai

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Sumit Roy, Chair

John Sahr

Jenq-Neng Hwang

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Compressive Detection and Estimation with Applications to Cognitive Radio and Radar

Linda Bai

Chair of the Supervisory Committee:
Professor Sumit Roy
Electrical Engineering

According to Nyquist Sampling theorem, a band-limited signal can be reconstructed accurately if the sampling rate exceeds twice the maximum frequency of the signal. In many scenarios, this Nyquist sampling rate cannot be achieved due to hardware limitations. Compressive sensing (CS) is a technique to reconstruct a signal from sub-Nyquist samples, given that the signal is sparse in a known domain [13][24]. The CS technique has been applied to different areas in the field of communications and networking. Of fundamental importance to current research is the need to adapt CS according to different requirements and constraints in each area.

In Chapter.2 and Chapter. 3, the application of compressive sensing to spectrum sensing in cognitive radio is discussed. Fast and reliable detection of available channels (i.e. temporarily unoccupied by primary users) is a *fundamental* challenge in cognitive radio design. The (mean) time to detection of an idle channel is governed by the (increasing) front-end *bandwidths* to be searched for a given (channel) bandwidth resolution. Wideband RF front-ends followed by suitable channelization and digital signal processing algorithms are consistent with speedier detection, but they also imply the need for very high speed analog-to-digital converters (ADCs) that are currently impractical. On the other hand, traditional heterodyne receiver architectures that consist of analog bandpass filtering require much lower rate ADCs, but at the expense of significant scanning operation steps, that constitute a roadblock towards lowering of the scan duration.

In Chapter.2, for detection of one idle channel, we propose a *multi-resolution spectrum sensing* scheme based on the principle of under-sampling (bandpass sampling) that provides a suitable middle ground between the above choices, i.e. our approach requires modest ADC sampling rate yet achieves fast scanning. A performance model for this architecture is developed based on an analysis for the mean time to detect an idle channel. The detection threshold and the sub-sampling factor are optimized jointly to minimize the mean detection time.

In Chapter.3, we propose a new *compressive spectrum sensing architecture based on bandpass sampling)* to detect all the idle channels in the given spectrum. Compared to other compressive spectrum sensing architectures, the proposed method does *not* require a high-speed Nyquist rate analog component. Numerical results show that the proposed schemes in these two chapters provide significantly faster idle channel detection than the conventional serial search scheme with a heterodyne architecture.

In Chapter.4, the application of compressive sensing to clutter subspace estimation in cognitive radar is discussed. Space-Time Adaptive Processing (STAP) based on matched filter processing in the presence of additive clutter (modeled as colored noise) requires knowledge of the clutter covariance matrix. In practice, this is estimated via the sample covariance matrix using samples from the neighboring range bins around the reference bin. By applying compressive sensing, the number of training samples needed to estimate the covariance matrix can be significantly reduced, provided that the basis mismatch problem, inherent to compressive sensing can be mitigated. This chapter presents an *adaptive* approach to choosing the best sparsifying basis, using dictionary learning to estimate the radar clutter subspace. Numerical results show that the proposed algorithm achieves the desired reduction in training samples, and is more accurate than previous reduced-rank algorithm baseline.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

ADC: Analog-to-Digital Converter

CGPAMF: Conjugate Gradient Parametric Adaptive Matched Filter

CS: Compressive Sensing

CSDL: Compressive Sensing with Dictionary Learning

CR: Cognitive Radio

DSA: Dynamic Spectrum Access

IB: Idle Block

IID: Independent and Identically Distributed

KASSPER: Knowledge Aided Sensor Signal Processing and Expert Reasoning

LO: Local Oscillator

LPF: Low-pass Filter

MWF: Multi-stage Wiener filter

OB: Occupied Block

PCI: Principal Component Inverse

P_D: Probability of Detection

P_{FA}: Probability of False Alarm

RF: Radio Frequency

RGS: Recursive Gram-Schmidt orthonormal basis selection algorithm

RIP: Restricted Isometry Property

RMB RULE: Reed-Mallett-Brennan rule

SINR: Signal-to-interference-plus-noise Ratio

SMI: Sample Matrix Inversion

SNR: Signal-to-Noise Ratio

STAP: Space-Time Adaptive Processing

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to everyone who has kindly offered help during my Ph.D. study.

First of all, I would like to express my heartfelt gratitude to my Ph.D. advisor, Prof. Sumit Roy. His immense knowledge and insightful guidance are always leading me towards the right direction. His ideas have supported me during the most difficult moment in my research and study. It is my pleasure to have him as my advisor.

In addition, I want to thank all the current and past members of my Ph.D. committee - Prof. Jenq-Neng Hwang, Prof. John Sahr, Prof. Maryam Fazel, Prof. Charles Barrack, and Prof. Barbara Henry, for their generous help and valuable suggestions.

Furthermore, I would like to thank all my UW labmates and friends who have helped me with my research and study. Hamed Firooz, Chittabrata Ghosh, Ling Luo, Rohit Gupta, Fei Ye, Wei Shi, Joshua Stahl, Colby Boyer, Abdulmohsen Mutairi, Farzad Hessar, Yue Yang, Hossein Safavi, Chang Wook Kim, Shwan Ashrafi, Yoshikazu Nishida, Benjamin Morgan, De Meng, Zixia Hu, Zhiyong Chen, Daibashish Gangopadhyay, Julie Hu, Xing Li, Xiang Zou, Prof. Payman Arabshahi, Prof. Hui Liu, and Prof. James Ritcey, all supported my work in various ways. Also, I want to thank my co-workers and advisors in AFRL, Wright-Patterson Air Force Base - Prof. Muralidhar Rangaswamy, Dr. Jeffrey Sanders, and Jason Parker.

At last I want to sincerely thank my family - my parents Jing Bai and Xian Wu, my husband Shaowu Huang, and my son Edwin Huang. Without their support, I will not be able to finish this work.

DEDICATION

to my parents Jing Bai and Xian Wu

Chapter 1

INTRODUCTION

1.1 Compressive Sensing

According to Nyquist Sampling theorem, a band-limited signal can be reconstructed accurately if the sampling rate exceeds twice the maximum frequency of the signal. In many scenarios, this Nyquist sampling rate cannot be achieved due to hardware limitations. Compressive sensing (CS) is a technique to reconstruct a signal from sub-Nyquist samples, given that the signal is sparse in a known domain [13][24]. The CS technique has been applied to various areas and fields, including image processing, smart grid, and sensor networks.

Sparsity is the inherent property of many different kinds of signals. For example, in network tomography, where the task is to find congested links using limited number of end-hosts in a network, the number of congested links is small compared to the total number of links [5]. Therefore, denote a congested link as '1' and a normal link as '0', the status for all the links is a sparse vector, with only a few non-zero elements in it. Another example would be the well-known results that most images are sparse in the DCT domain, with only a few dominant coefficients. In this thesis, the underlying sparsity of the radio spectrum (especially $0.1 - 10$ GHz band) and the clutter ridge in radar signal processing is exploit, and compressive sensing is applied in both applications to improve performance.

1.2 Brief Review of Compressive Sensing Theory

Compressive sensing (CS) is a well-known technique to reconstruct a sparse N-dimensional signal vector $\mathbf{x}$ from an M-dimensional representation $\mathbf{y} = \mathbf{\Phi\Psi x}$ ($M \ll N$) [24], where the $N \times N$ matrix $\mathbf{\Psi}$ is the sparsifying basis. In the spectrum sensing scenario, $\mathbf{\Psi}$ is the Inverse Fourier Transform matrix, and $\mathbf{x}$ represents the amplitude spectra in the frequency domain. In the scenario of STAP algorithm, Φ is a matrix whose columns include the basis steering vectors of the clutter subspace, and Ψ is the identity matrix. The rows of the

matrix $\mathbf{\Phi}$ can be seen as the M basis vectors onto which the higher N-dimensional signal is projected to obtain the lower M-dimensional representation $\mathbf{y}$. There exists a mature literature regarding appropriate choices of the observation matrix $\mathbf{\Phi}$, whose elements are typically i.i.d Gaussian or Bernoulli variables [24]. An appropriate choice of $\mathbf{\Phi}$ must satisfy the Restricted Isometry Property (RIP), formally defined as follows [12],

Definition 1 ***Restricted Isometry Property*** *A matrix* $\mathbf{\Phi}$ *satisfies RIP with parameter* (k, δ_k) *if*

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{1.1}$$

for all k-sparse vectors $\mathbf{x}$*, where k-sparsity means there are at most k non-zero elements in the vector* $\mathbf{x}$*.*

Given $\mathbf{\Phi}$ satisfies RIP with $(2k, \delta_{2k})$ and $0 < \delta_{2k} < \sqrt{2} - 1$, the k-sparse vector $\mathbf{x}$ is reconstructed using

$$\begin{aligned} &minimize\ \|\mathbf{x}\|_1 \\ &subject\ to\ \mathbf{y} = \mathbf{\Phi\Psi x} \end{aligned} \tag{1.2}$$

1.3 Cognitive Radio

Continued proliferation of wireless services has led to renewed scrutiny of spectrum utilization. In particular, several licensed bands, e.g., land mobile and amateur radio, have been shown to be under-utilized with occupancy rates between 15% to 85% by several recent measurements of spectrum utilization from Federal Communications Commission (FCC) and National Telecommunications and Information Administration (NTIA)[49]. Cognitive radios - enabled by emerging software defined transceiver architectures - is an approach for enhanced spectrum utilization, by allowing unlicensed users to opportunistically access licensed bands, when the latter are unused. The unlicensed or secondary users must sense the spectrum to detect idle channels (free of incumbents or primary users) and are allowed to utilize the idle channels, provided they do not interfere with the licensed users, i.e., vacate the channel when a primary user transmission is detected.

Therefore, one of the primary requirements for cognitive radio is *fast and effective detection* of idle channels to exploit the available secondary network capacity [49]. Due to the large bandwidth (for example, up to 10 GHz) that needs to be sensed, speedy idle channel detection is a challenging task. Assume that the entire (one-sided) spectrum of bandwidth $W = N\,W_c$ is channelized into an N-set of contiguous frequency domain channels of bandwidth W_c each. The choice of receiver architecture is a major factor that influences the mean detection time of all the idle channels among these N channels. The heterodyne architecture is the most popular legacy architecture due to its high selectivity and sensitivity [1]. Heterodyne receivers use analog bandpass filters for selecting a W_c-width channel as shown in Fig.1.1. The output is then down-converted to baseband and fed into an analog-to-digital converter (ADC) [67]. The ADC samples at $f_s = W_c$ (to produce the complex baseband signal), at a rate which is much smaller than the total signal bandwidth $N\,W_c$. The receiver decides the status of the selected channel using one of several standard detection algorithms described in [70]. After the status of the selected channel is determined, the receiver switches to the next channel in sequence. A conventional serial search scheme (implying a sweep or step-through of the LO frequency) is used to determine the occupancy status of the N channels. Depending on the bandwidth needs required by the applications, secondary users may seek to detect one idle channel or *all* available idle channels. For detection of one idle channel, the receiver stops searching as soon as it detects the first idle channel, whereas in the latter, the receiver will search the status of all channels in the spectrum of interest. In Chapter 2, we focus on the first scenario; while in Chapter 3, we focus on the second, consistent with the desire to maximize secondary throughput.

For the heterodyne receiver architecture, the time to detect idle channels has two components - a) the integration time T_i required by the detection algorithm once a channel has been selected for investigation; and b) the duration for switching to a new channel T_{sw} during scanning, that is largely determined by the design/operation of the LO [47]. Denote N_i as the number of samples required for meeting desired detection accuracy in each channel for a given detection algorithm, then $T_i \approx N_i/W_c$. The mean detection time for the heterodyne receiver to determine status for *all* channels is thus $T_{d,serial} = N(T_i + T_{sw})$. For example, for $W_c = 200$ kHz, $N = 1024$, $N_i = 1000$, and $T_{sw} \approx 50\mu s$ as in [63], then

$T_i = 5$ ms, and $T_{d,serial} \approx \mathbf{5.17}$ sec.

On the other hand, in homodyne receiver architectures such as that in Fig.1.1, the wideband RF received signal is sampled directly by an ADC, followed by subsequent digital signal processing for channelization and determination of channel status. For digital signal processing, the power spectrum is obtained by squared magnitude of the N-point FFT for the wideband signal. Each bin in the power spectrum represents the estimated power in the corresponding channel, and is compared with a power threshold to determine the status of the channel [11].

An estimate of the (mean) time to determine status of all N channels is obtained by noting that the total number of samples needed is $N \cdot N_i$, when N_i samples are required for each channel. Since the sampling rate of the ADC for the complex baseband signal is $N \cdot W_c$, the acquisition time is $(N \cdot N_i)/(N \cdot W_c) = N_i/W_c$. Note that signals in all the N channels are acquired in parallel and hence the analog circuit switching time T_{sw} is avoided, yielding an acquisition time that is independent of N. Denote T_{dsp} as the latency due to the subsequent signal processing operations such as the FFT in generating the spectrogram. Hence the time to determine status of all channels for such a homodyne architecture is $T_{d,sdr} = N_i/W_c + T_{dsp}$. For the spectrum sensing task above, the detection time for all the idle channels with the software defined radio architecture (i.e., the homodyne architecture) is approximately **31** ms, with $T_{dsp} = N_i \cdot 26\mu s$, where $26\mu s$ is the typical processing time for a 1024-point FFT [22]. Thus, software defined radio (SDR) based transceivers can potentially achieve faster detection by two orders of magnitude, at the expense of an ultra-fast (multi-GHz) ADC for Nyquist sampling of the wideband signal, that considerably increases the power consumption and IC complexity [4].

1.3.1 Application of Compressive Sensing in Cognitive Radio

A snapshot of the power spectrum based on measurements taken in Aachen, Germany, is shown in Fig.1.2 [84]. With a threshold of -111 dBm as discussed in Chapter 2, it is observed that spectrum occupancy (availability) in licensed bands shows significant variability as a function of frequency. For instance, at a given location, for some part of the spectrum, most

channels may be occupied while in another spectral segment, most channels may be idle [31]. The spectrum occupancy/idle status can be represented by a vector with binary entries, where each 0(1) represents the status of a channel (free or busy). Compressive sensing (CS) is a well-known technique for recovery of *sparse* signals such as the above binary vector, where sparsity implies that *the l_0 norm (number of non-zero entries) is small.* Denote the average probability of a channel being idle as $\bar{p}$, then a sparse spectrum indicates $\bar{p} \approx 1$.

When the power spectrum of the incoming signal for the receiver is therefore a sparse vector, with elements representing the power in each of the N channels, compressive sensing can be applied to reduce the ADC sampling rate for the homodyne architecture. In Chapter 3, we unveil a *new, compressive* architecture based on sub-Nyquist [1] sampling ADCs. Our proposal thus treads a middle ground by combining the strengths of the two architectures, in terms of achieving fast detection, while requiring a lower rate ADC.

1.4 Space Time Adaptive Processing

Space-Time Adaptive Processing (STAP) is a widely used radar signal processing technique to detect a multi-channel signal of interest (target) in temporally and spatially correlated clutter [50]. Due to motion of the aerial platform, the target cannot be efficiently detected from the clutter background by using spatial or temporal information alone, leading to the necessity of space-time joint processing. A matched filter to suppress clutter using it's statistics is fundamental to effective STAP. To construct the matched filter, an estimate of the clutter covariance matrix is needed, that usually requires numerous homogenous samples from neighboring range cells. This is generally infeasible, consequently there is a rich literature on reducing the number of samples by exploiting prior information [85][33]. Many of these techniques involves exploiting the low-rank nature of the covariance matrix, such as [32] and [42].

---

[1] The sampling is sub-Nyquist relative to the rate required to sample the (bandpass) spectra treated as a low pass signal.

1.5 Application of Compressive Sensing in Cognitive Radar

Compressive sensing (CS) is a well-known technique that has rapidly risen to prominence in the signal processing communities for it's ability to provide innovative solutions to problems of signal estimation, where the signal of interest is inherently *sparse* in some basis. For a vector signal in STAP, this is equivalent to exploiting the low-rank nature of the clutter subspace in the angular and (temporal) Doppler frequency domains that has the potential to significantly reduce the required number of samples (i.e., neighboring range bins) required for estimation of the clutter covariance matrix [72], [56] . However, in [72], the angle and frequency axis are discretized on a uniform grid, via the use of steering vectors corresponding to the lattice points as the sparsifying basis. As clutter patches may not be precisely located at the grid points, the resulting basis mismatch can lead to inaccurate clutter subspace estimation [17], in turn leading to significant degradation in STAP performance.

In Chapter 4, we propose the use of dictionary learning algorithm to estimate the clutter subspace, where the angle and frequency axis are discretized *non-uniformly and adaptively* to mitigate the basis mismatch and obtain more accurate estimates of the clutter covariance matrix.

1.6 Dissertation Organization

We first brieftly review the theory and algorithms of CS, which are used in the following chapters. In Chapter 2, the application of compressive sensing to spectrum sensing in cognitive radio to detect one idle channel is discussed; while in Chapter 3, CS is applied to detect all the idle channels using bandpass sampling. In Chapter 4, the application of CS in cognitive radar is discussed, and dictionary learning and CS are combined to improve the detection performance of STAP. Finally, in Chapter 5, we conclude this dissertation.

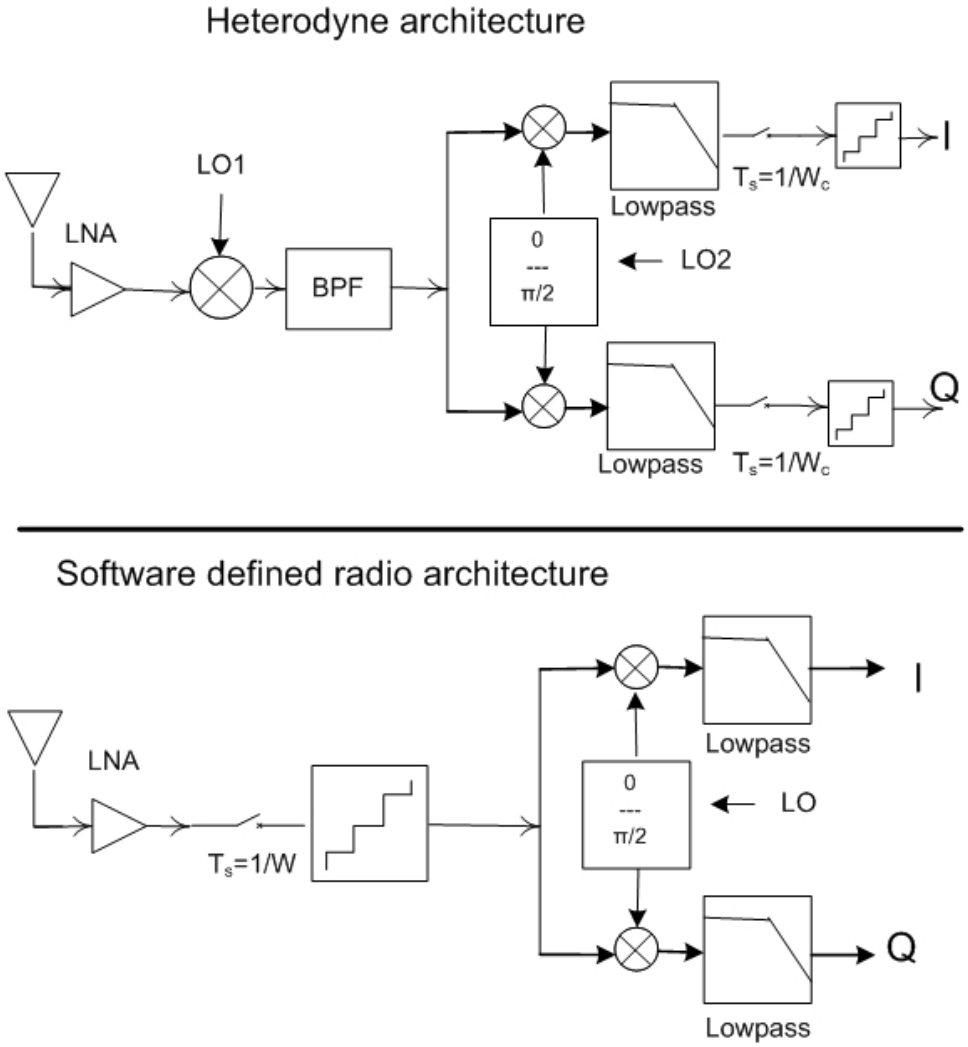


Figure 1.1: Block diagram of the heterodyne and homodyne receiver architectures

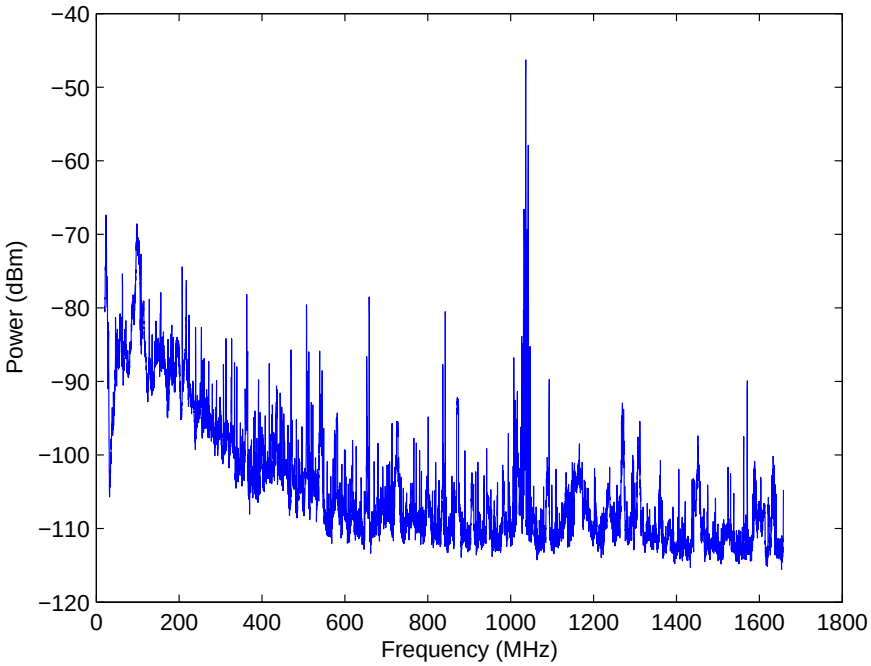


Figure 1.2: A snapshot of the power spectrum from 20MHz to 1.64GHz, taken in Aachen, Germany

Chapter 2

MULTI-RESOLUTION IDLE CHANNEL DETECTION WITH BANDPASS SAMPLING

2.1 Introduction

In this chapter, we seek to find an idle channel in the radio spectrum. As soon as an idle channel is found, the receiver stops searching. Assume that the entire spectrum is channelized into an N-set of contiguous (frequency domain) channels, each with bandwidth W_c; each of them is idle at some given instant with a probability of $p = \frac{L}{N}$.

The mean detection time to find an idle channel is influenced by the choice of receiver architecture. The heterodyne architecture is the most popular architecture due to its high selectivity and sensitivity [1].

Recall that in Chapter 1, we showed that, using the heterodyne receiver architecture, the time to detect an idle channel has two components - a) the integration time required by the detection algorithm once a channel has been selected for interrogation T_i and b) the duration for switching to a new channel T_{sw} [47]. The mean detection time for one-stage serial search is given by [48]

$$T_{d,serial} = \frac{N}{L+1}(T_i + T_{sw}) \tag{2.1}$$

To reduce the mean detection time, a two-stage sensing scheme was proposed in [47]. The total spectrum is partitioned into several blocks of equal size. Each block contains α channels. In the first stage, the receiver searches the entire block for idle channel(s). Based on the outcome of the first stage, a successful block (i.e. one that indicates the presence of idle channels) is searched to locate an individual idle channel by a serial search within the block. As can be expected, traditional approaches - including one-stage and two stage schemes - suffer from increased (mean) time to detection as p decreases.

As discussed in Chapter 1, current transceiver architectures must address the funda-

mental dichotomy of the trade-off between sensing speed and ADC sampling rate. In this chapter, we unveil a *new, multi-resolution* approach based on sub-Nyquist [1]sampling rate ADC that treads a middle ground, in terms of achieving fast detection, while requiring a lower rate ADC.

2.1.1 Sub-Nyquist Sampling Architecture based on Compressive Sampling Theory

It is observed that the spectrum occupancy (availability) in licensed bands shows significant variability as a function of geographic location and frequency. For instance, at a given location, some part of the spectrum most channels may be occupied while in another spectral segment, most channels may be idle. The spectrum occupancy/idle status can be represented by a vector with binary entries, where each 0(1) represents the status of a channel (free/busy); the prior discussion suggests that in many scenarios, such a vector may be 'sparse', i.e. a significant proportion of 0's or 1's.

Compressive sampling (CS) is a well-known technique for recovery of such sparse signal $\mathbf{X}$ (N-dimensional vector) from a succinct approximate representation $\mathbf{\Phi\Psi X}$ (M-dimensional vector, where $M \ll N$) [24]. It should be noted that a majority of the existing CS literature have focussed on the *signal reconstruction* problem. Various receiver architectures with sub-Nyquist sampling for this problem have been proposed. In [39], [45], [54], and [77], a demodulator multiplies the received (analog) signal with a random Bernoulli chipping sequence; the resulting product is integrated over a symbol duration and sampled at a sub-Nyquist rate. In [28], a non-uniform sampler is used to generate sub-Nyquist samples for reconstruction. Although the sampler in the ADC operates at sub-Nyquist rate, the architecture requires fast (Nyquist rate) analog chipping sequence generators and multiplier.

Applications of CS to the problem of spectral detection/estimation have been limited, beginning with [75]. In [20] and [66], the spectrum is reconstructed with high accuracy using l_1 minimization algorithm, which results in accurate detection in the following step. In [64], the auto-correlation of the sub-Nyquist samples was used for spectrum estimation, that reduced the required SNR. It should be noted that although traditional CS algorithms,

[1]The sampling is sub-Nyquist relative to the rate required to sample the (bandpass) spectra treated as a low pass signal.

such as l_1 minimization, work only for the scenario where $p \sim 0$ (most of the channels are idle), greedy pursuit algorithm with maximum likelihood detection can detect idle channels when $p \sim 0$ and $p \sim 1$. The work closest to ours in spirit is [60], that uses an undersampled A/D converter on the incoming FM signal and reconstructs using a quadrature mirror filter.

Nevertheless, these approaches based on CS can provide accurate results only if most of the channels in the spectrum are in the same status (idle/occupied). If the number of occupied channels is close to the number of idle channels, the CS-based compressive ADCs provides inaccurate estimates [13], which leads to low P_d and high P_{fa}. Also, sub-Nyquist sampling worsens the SNR of the signal in reconstruction, due to aliasing of noise.

2.1.2 Contributions

Our work is a synthesis of the notion of Compressive Sensing within a bandpass sampling architecture. The receive chain with bandpass sampling is shown in Fig.1.1 where the signal $r(t)$ is sampled at a sub-Nyquist rate $f_s = \frac{1}{T_s} = \frac{NW_c}{\gamma}$ to obtain $y[n] = r(nT_s)$, where T_s is the sampling period, and $\gamma > 1$ is the sub-sampling factor. $y[n]$ is then passed through a quantizer and fed into the digital signal processing block. This architecture in RF receivers [81] achieves frequency down-conversion by undersampling $r(t)$ (viewed as a wideband lowpass signal) [34], where the sampling rate is determined by the bandwidth of the desired bandpass signal (i.e. conforms to the bandpass Nyquist sampling theorem) instead of the carrier frequency [3][86] (corresponding to the lowpass Nyquist sampling theorem). Owing to the reduction of the sampling rate, bandpass sampling simplifies the design of the local oscillator, leading to better performance than a RF mixer for down-conversion [67].

In Sec.4.2, a spectrum sensing scheme with bandpass sampling is proposed. Here, the entire spectrum (in principle) can be sensed simultaneously, obviating the need for circuit switching. However, bandpass sampling technique worsens the effective SNR for detection, due to noise aliasing [67]. Therefore, our work espouses a multi-resolution sensing approach [47] that consists of a preliminary coarse resolution sensing (CRS) with bandpass sampling followed by fine resolution sensing (FRS) with serial search, an architecture initially

developed in [47].

We use the mean time to detect a free channel as the figure of merit for a sensing algorithm subject to performance guarantees - notably that the probability of detection and false alarm be in a reasonable range: $P_d > 0.9, P_{fa} < 0.1$ [10]. A high false alarm rate will cause interference to primary users, while a high detection rate is necessary for a high spectral utilization rate. Thus, we seek to minimize the mean detection time, subject to the constraints on P_d and P_{fa}. Two design parameters - the sub-sampling factor and the threshold in energy detection are optimized jointly. In Sec.4.4, numerical results for the optimized mean detection time are developed for the proposed multi-resolution scheme. A notable feature is that our multi-resolution scheme achieves *consistent performance for all* $p \in (0, 1]$, whereas the traditional one-stage sensing scheme have poor performance for small p, and the compressed sensing based schemes have poor performance for $p \sim 0.5$. Impact of channel occupancy rate and SNR on the optimal design parameters is also explored. Finally, performance comparison of the one-stage serial search, conventional two-stage, and the proposed scheme based on a real world spectrum dataset obtained in Aachen, Germany is also presented, for further validation.

2.2 System Description

2.2.1 System Model

Assume the central frequency of the i^{th} channel is $f_i = iW_c - \frac{W_c}{2}$, where W_c is the bandwidth for each channel. A binary variable, O_i is used to denote the status of the i^{th} channel, where $O_i = 0(1)$ means the channel is busy (idle). We will invoke an i.i.d Bernoulli model for channel status, i.e.,

$$Pr(O_i = 1) = p, i = 1, ..., N \tag{2.2}$$

where the signal power in the i^{th} channel is also a Bernoulli variable, i.e.,

$$P_{s,i} = P_s(1 - O_i) \tag{2.3}$$

2.2.2 Bandpass Sampling

Denote the received signal as $r(t)$ whose bandwidth $W = N\,W_c$. The discrete-time Fourier transform $Y(f)$ of the signal $y[n]$ sampled at the rate f_s is related to $R(f)$, the Fourier transform of $r(t)$, by the well-known result

$$Y(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} R(f - kf_s) \tag{2.4}$$

An example is illustrated in Fig.2.1 where the *complex baseband* signal $r(t)$, comprising of 12 channels each of bandwidth W_c, is sampled at a (sub-Nyquist) rate of $f_s = 5W_c$. As shown in Fig.2.1, aliasing occurs in the spectra of the sampled signal due to the sub-Nyquist sampling rate; the sixth channel folds back onto the first channel, since $f_6 - f_s = \frac{11}{2}W_c - 5W_c = \frac{1}{2}W_c = f_1$. Clearly, when a signal is sub-sampled with rate f_s, all frequency components $\{f_i \pm kf_s\},\ k = 0, 1, \ldots$ present in the signal alias back onto the same frequency $f_i \in [0, f_s]$, where $[0, f_s]$ is the first Nyquist zone. Letting $f_s = \frac{N}{\gamma}W_c$, where $\frac{N}{\gamma}$ is an integer, the relationship between $Y_i,\ i = 1, ..., \frac{N}{\gamma}$, (the Fourier coefficients of $y[n]$ on the interval $[0, f_s]$) and $R_i,\ i = 1, ..., N$, (the Fourier coefficients of $r[n]$ sampled at Nyquist rate NW_c on the interval $[0, NW_c]$) is

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{\frac{N}{\gamma}} \end{bmatrix} = \frac{1}{\gamma}\mathbf{\Phi} \begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix} = \frac{1}{\gamma}\mathbf{\Phi}\mathbf{R} \tag{2.5}$$

where the $\frac{N}{\gamma} \times N$ matrix

$$\mathbf{\Phi} = \begin{bmatrix} \overbrace{1\ 0\ 0\ ...\ 0}^{\left(\frac{N}{\gamma}-1\right)0's} & 1 & 0 & 0 & ... & 0 & 1 & 0 & 0 & ... \\ 0\ 1\ 0\ ...\ 0 & 0 & 1 & 0 & ... & 0 & 0 & 1 & 0 & ... \\ 0\ 0\ 1\ ...\ 0 & 0 & 0 & 1 & ... & 0 & 0 & 0 & 1 & ... \\ & & & & \vdots & & & & & \\ 0\ 0\ 0\ ...\ 1 & 0 & 0 & 0 & ... & 1 & 0 & 0 & 0 & ... \end{bmatrix} \tag{2.6}$$

For example, let $N = 10$, and $\gamma = 2$. Then

$$\begin{aligned} Y_1 &= \sum_{i=0}^{4} r(i/f_s) = \sum_{i=0}^{4} r(i\gamma/NW_c) \\ &= \sum_{i=0}^{4} r(2i/NW_c) \end{aligned} \tag{2.7}$$

Since

$$R_1 = \sum_{i=0}^{9} r(i/NW_c) \tag{2.8}$$

and

$$\begin{aligned} R_5 &= \sum_{i=0}^{9} r(i/NW_c) e^{\frac{j2\pi}{10}5i} \\ &= \sum_{i=0}^{4} r(2i/NW_c) - \sum_{i=0}^{4} r((2i+1)/NW_c) \end{aligned} \tag{2.9}$$

it follows that $Y_1 = \frac{1}{2}(R_1 + R_5)$.

Each element in $\mathbf{Y}$ is thus determined by $\lfloor\gamma\rfloor$ or $\lfloor\gamma\rfloor + 1$ elements in $\mathbf{R}$, corresponding to a subset (block) of channels [47]. As shown in Fig.2.1, by comparing each element in $\mathbf{Y}$ to a threshold δ, each block can be classified as either an Idle Block (IB) or a Occupied Block (OB) without any idle channels, respectively. A binary variable X is used to denote the status of a block, where $X = 0(1)$ means the block is occupied or idle. In the following, we assume γ is an integer that divides N.

2.2.3 Multi-resolution Sensing

Secondary users search for idle channels using the multi-resolution (two-stage) sensing approach developed in [47], with minor modifications, as depicted in Fig.2.2. Our scheme differs from the original [47] primarily in Stage I (i.e. coarse sensing stage): we sense all the blocks and locate all the IBs simultaneously with bandpass sampling, whereas the original [47] performs coarse resolution sensing sequentially and thus locates (at most) one IB (the first to be detected) and stops. The effect of this difference on the overall sensing time is discussed in Sec.4.3. In Stage II, the detected IB is subjected to a fine-grained serial search to identify an idle channel. If all channels in this IB turn out to be occupied, the receiver

resorts to the next IB in sequence and repeats. When all the IBs in the entire spectrum have been searched and yet no idle channel is found, the receiver goes back to Stage I again to obtain new time samples and repeats the process.

In stage I, the energy in the entire block is compared to a threshold δ to estimate status, i.e. $X = 1$ if $P_{block} < \delta$, corresponding to one of the two hypotheses H_1 ($X = 1$) or H_0 ($X = 0$). From Eq.(3.2),

$$Pr(H_0) = (1-p)^{\gamma} \tag{2.10}$$

$$Pr(H_1) = 1 - (1-p)^{\gamma} \tag{2.11}$$

Let $H_{1.k}$ denote the event where there are exactly k idle channels in a block. Then,

$$Pr(H_{1,k}) = \left(\begin{array}{c} \gamma \\ k \end{array} \right) p^k (1-p)^{\gamma-k}, \quad k = 1, 2, .., \gamma \tag{2.12}$$

Since the presence/absence of signals in different channels are independent, the R_i are independent variables. For symmetric signal constellations,

$$E(R_i) = 0 \tag{2.13}$$

From (3.4), the expectation of the signal power in a block is

$$\begin{aligned} \overline{P}_{block,signal} &= E(|\sum_{i=1}^{\gamma} R_i|^2) = E(\sum_{i=1}^{\gamma} R_i)(\sum_{i=1}^{\gamma} R_i^*) \\ &= E(\sum_{i=1}^{\gamma} |R_i|^2 + \sum_{i \neq j} R_i R_j^*) = \sum_{i=1}^{\gamma} E(|R_i|^2) + \sum_{i \neq j} E(R_i)E(R_j) \\ &= \sum_{i=1}^{\gamma} E(|R_i|^2) = \sum_{i=1}^{\gamma} P_{s,i}\,, \end{aligned} \tag{2.14}$$

where $P_{s,i}$ is the signal power in the ith channel.

To improve detection, M_T samples of the measured Fourier amplitude of the corresponding block $\overline{Y}_{block}$ are used to determine the energy in a block [11], i.e. the energy in a block

$\overline{P}_{block} = \sum_{M_T} |\overline{Y}_{block}|^2$. $\overline{P}_{block}$ follows a non-central chi-square distribution with a degrees of freedom $2M_T$ [2]. The probability of detection for an idle channel in the first stage is thus

$$\begin{aligned} P_d(\delta,\gamma) &= Pr(\overline{P}_{block} < \delta | H_1) = \frac{Pr(\overline{P}_{block} < \delta, H_1)}{Pr(H_1)} \\ &= \frac{\sum_{k=1}^{\gamma} Pr(H_{1,k}, \overline{P}_{block} < \delta)}{Pr(H_1)} \\ &= \frac{1}{Pr(H_1)} \sum_{k=1}^{\gamma} Pr(\overline{P}_{block} < \delta | H_{1,k}) Pr(H_{1,k}) \\ &= \frac{1}{1-(1-p)^{\gamma}} \sum_{k=1}^{\gamma} [1 - Q_{\frac{M_T}{2}}(\sqrt{\frac{M_T(\gamma-k)P_s}{\gamma\sigma_n^2}}, \\ &\quad \sqrt{\frac{\delta}{\gamma\sigma_n^2}})] \begin{pmatrix} \gamma \\ k \end{pmatrix} p^k (1-p)^{\gamma-k} \end{aligned} \tag{2.15}$$

where $Q_{x_1}(x_2, x_3)$ is the Marcum Q function [2]. The probability of false alarm in the first stage is

$$\begin{aligned} P_{fa}(\delta,\gamma) &= Pr(\overline{P}_{block} < \delta | H_0) \\ &= 1 - Q_{\frac{M_T}{2}}(\sqrt{\frac{M_T P_s}{\sigma_n^2}}, \sqrt{\frac{\delta}{\gamma\sigma_n^2}}) \end{aligned} \tag{2.16}$$

Denote the average circuit switching time between channel sensing as T_{sw} (approximately $30\mu s$ based on current circuit designs [18]), and the average integration time for sensing a channel as T_i (about $1\mu s$ for medium SNR environment (SNR$\sim$ 10 dB) and $M_T \approx 30$ [47]). Consider the following idealized assumption for the second stage: probability of detection $= 1$, and the probability of false alarm $= 0$. The average number of switching and integration steps required when $X = 1$ is

$$\begin{aligned}
\overline{S}_{IB} &= E(number\ of\ switching|X=1) \\
&= 1 \times Pr(the\ first\ channel\ idle|X=1) \\
&\quad +2 \times Pr(the\ first\ channel\ busy,\ the\ second\ channel\ idle|X=1)+... \\
&\quad +\gamma \times Pr(the\ first\ \gamma-1\ channels\ busy,\ the\ last\ channel\ idle|X=1) \\
&= 1 \times \frac{p}{1-(1-p)^{\gamma}} + 2 \times \frac{(1-p)p}{1-(1-p)^{\gamma}} + ... + \gamma \times \frac{(1-p)^{\gamma-1}p}{1-(1-p)^{\gamma}} \\
&= \frac{1}{p} - \gamma \frac{(1-p)^{\gamma}}{1-(1-p)^{\gamma}} \qquad (2.17)
\end{aligned}$$

Hence the mean duration till finding an idle channel when $X = 1$ with serial search scheme is

$$\overline{T}_{IB} = (\frac{1}{p} - \gamma \frac{(1-p)^{\gamma}}{1-(1-p)^{\gamma}})(T_i + T_{sw}) \qquad (2.18)$$

Let $T_{c,i}$ be the net integration plus switching time spent on the i^{th} block in the spectrum in the second stage. The average time for the second stage is then

$$\begin{aligned}
\overline{T}_{stageII} &= E(\sum_{i=1}^{\frac{N}{\gamma}} T_{c,i}) = \sum_{i=1}^{\frac{N}{\gamma}} E(T_{c,i}) \\
&= (Pr(H_1)P_d\overline{T}_{IB} + Pr(H_0)P_{fa}(T_i + T_{sw})\gamma) \\
&\quad +((1-Pr(H_1)P_d)(Pr(H_1)P_d\overline{T}_{IB} + Pr(H_0)P_{fa}(T_i + T_{sw})\gamma) + ... \\
&\quad +((1-Pr(H_1)P_d)^{\frac{N}{\gamma}-1}(Pr(H_1)P_d\overline{T}_{IB} + Pr(H_0)P_{fa}(T_i + T_{sw})\gamma)) \\
&= \frac{1}{Pr(H_1)P_d}(Pr(H_1)P_d(\frac{1}{p} - \gamma \frac{(1-p)^{\gamma}}{1-(1-p)^{\gamma}}) \\
&\quad +Pr(H_0)P_{fa}\gamma)(T_i + T_{sw})(1-(1-Pr(H_1)P_d)^{\frac{N}{\gamma}}) \qquad (2.19)
\end{aligned}$$

The average number of Stage-1 detection cycles is given by

$$\overline{S} = \frac{1}{1-(1-Pr(H_1)P_d)^{\frac{N}{\gamma}}} \qquad (2.20)$$

Hence the mean detection time $\overline{T}_d$ - now explicitly recognized as a function of δ and γ - is

$$\overline{T}_d(\delta,\gamma) = \overline{S}(\overline{T}_{stageI} + \overline{T}_{stageII}) \tag{2.21}$$

where the average time spent in Stage-1 [11] is proportional to the bandwidth to be sensed. Therefore,

$$\overline{T}_{stageI} = \frac{N}{\gamma} T_i \tag{2.22}$$

For each value of the channel occupancy rate p, the optimal δ and γ are found by

$$\begin{gathered} argmin\ \overline{T}_d(\delta,\gamma) \\ subject\ to \\ P_d(\delta,\gamma) \geq 0.9 \quad P_{fa}(\delta,\gamma) \leq 0.1 \end{gathered} \tag{2.23}$$

2.3 Minimizing Mean Detection Time

2.3.1 Optimal Threshold and Sub-sampling Factor

The algorithm to obtain optimal threshold and sub-sampling factor jointly is shown in Table.2.1. Because the sub-sampling factor γ belongs to the integer set $1 \leq \gamma \leq N \leq 1000$ in many practical applications, it is feasible to use exhaustive search with respect to γ in (2.23). In each iteration, an optimal threshold δ is calculated for a given γ, and the jointly optimal δ and γ obtained as a result.

The golden section algorithm [8] is applied to find $\hat{\delta} = \arg\min_{\delta} \overline{T}_d(\delta,\gamma)$ in Table.2.1. From (2.15) and (2.16), both P_d and P_{fa} increase monotonically with respect to δ. Therefore, for a fixed $\gamma = \gamma_{fix}$, Eq. 2.23 becomes

$$\begin{gathered} argmin\ \overline{T}_d(\delta,\gamma_{fix}) \\ subject\ to \\ \delta_{min} \leq \delta \leq \delta_{max} \end{gathered} \tag{2.24}$$

where $P_d(\delta_{min},\gamma_{fix}) = 0.9$, and $P_{fa}(\delta_{max},\gamma_{fix}) = 0.1$. δ_{min} and δ_{max} are obtained via an iterative binary search algorithm shown in Table. 2.2. First, an initial interval of $[\delta_{low},\delta_{high}]$

Initialization
$T(\gamma) = 0, \gamma = 1, ..., N$
Iterations
for $\gamma = 1 : N$
$\hat{\delta} = \arg\min_{\delta} \overline{T}_d(\delta, \gamma)$
$T(\gamma) = T_d(\hat{\delta}, \gamma)$
Output $min(\mathbf{T})$

Table 2.1: Pseudo code for brute force search

is chosen. Then, if the difference between $P_d((\delta_{low} + \delta_{high})/2, \gamma_{fix})$ and 0.9 is smaller than ϵ ($\epsilon \ll 0.9$), set $\delta_{min} = (\delta_{low} + \delta_{high})/2$. Otherwise, if the above condition is not satisfied, set $\delta_{high} = (\delta_{low} + \delta_{high})/2$ ($\delta_{low} = (\delta_{low} + \delta_{high})/2$) if $P_d((\delta_{low} + \delta_{high})/2, \gamma_{fix}) > 0.9$ ($P_d((\delta_{low} + \delta_{high})/2, \gamma_{fix}) < 0.9$). The above follows because $P_d(\delta, \gamma_{fix})$ increases monotonically with respect to δ and the algorithm iterates to update δ_{high} and δ_{low}. A similar recursion estimates δ_{max}.

Because $\overline{T}_d(\delta, \gamma_{fix})$ is a continuous smooth function over $\delta \in [\delta_{min}, \delta_{max}]$, $\exists \hat{\delta} \in [\delta_{min}, \delta_{max}]$, s.t.$\overline{T}_d(\hat{\delta}, \gamma_{fix}) \leq T_d(\delta, \gamma_{fix})$, $\forall \delta \in [\delta_{min}, \delta_{max}]$. To solve for $\hat{\delta}$ requires finding the minimum of a single-variable function on a fixed interval, which can be readily achieved via the golden section search and parabolic interpolation algorithm in Matlab [8].

2.3.2 Lower bound for Mean Detection Time

For a given γ, a lower bound for mean detection time of the multi-resolution sensing scheme with bandpass sampling can be found when $P_d = 1$ and $P_{fa} = 0$ (i.e. perfect detection in the first stage). Assume there exists at least one idle channel in the spectrum. From (2.21),

$$\overline{T}_d \geq \frac{N}{\gamma} T_i + (\frac{1}{p} - \gamma \frac{(1-p)^\gamma}{1-(1-p)^\gamma})(T_i + T_{sw}) \tag{2.25}$$

Similarly, a lower bound of mean detection time for the original two-stage scheme in [47] with bandwidth factor α is

Initialization
$\delta_{low} = 0$
$\delta_{high} = 1000$
$\delta = 0$
Iterations
do
$\delta = (\delta_{low} + \delta_{high})/2$
if $P_d(\delta, \gamma_{fix}) > 0.9$
$\delta_{high} = \delta$
else if $P_d(\delta, \gamma_{fix}) < 0.9$
$\delta_{low} = \delta$
while $\|P_d(\delta, \gamma_{fix}) - 0.9\| > \epsilon$
Output $\delta_{min} = \delta$

Table 2.2: The binary search algorithm

$$\overline{T}_{d,original} \geq (\frac{1}{1-(1-p)^{\alpha}} + (\frac{1}{p} - \alpha\frac{(1-p)^{\alpha}}{1-(1-p)^{\alpha}}))(T_i + T_{sw}) \tag{2.26}$$

Therefore, if $\alpha = \gamma$ (the same block size for both schemes), under the condition of perfect detection in both stages,

$$\overline{T}_{d,perfect} \leq \overline{T}_{d,original,perfect} \Leftrightarrow \frac{N}{\gamma}T_i \leq \frac{1}{1-(1-p)^{\gamma}}(T_i + T_{sw}) \tag{2.27}$$

Because in current circuits designs, $T_i \ll T_{sw}$ [48], the inequality in (2.27) is usually valid. Hence the mean detection time for the multi-resolution scheme is proven to be smaller than the original two-stage scheme, attributable mostly to the savings achieved in the coarse resolution sensing stage wherein all the blocks are sensed in parallel, avoiding significant circuit switching time.

2.4 Numerical Results

In this section, we first compare the performance of the proposed multi-resolution scheme with one-stage serial search, random search, and the original two-stage scheme in [47]. Then the impact of SNR and occupancy rate p on the mean detection time and the optimal δ and γ is discussed, with $T_{sw} = 30\mu s$, $T_i = 1\mu s$, $M_T = 50$, and $N = 100$. For the cases where γ does not divide N, $\frac{N}{\gamma}$ is substituted by $\lceil \frac{N}{\gamma} \rceil$.

2.4.1 Mean Detection Time

Fig.2.3 illustrates the mean detection time for one-stage random search, the original two-stage scheme, and the proposed multi-resolution scheme with bandpass sampling at SNR = 10 dB. Note that the mean detection time is similar for random search and serial search [48]. The optimal γ is 3 when $p \leq 0.45$, and 4 when $p > 0.45$. The results show that the multi-resolution scheme consistently outperforms the original two-stage scheme for $p \in (0, 0.5)$. For low p, the improvement in mean detection time is about 80% compared with one-stage search, and about 50% compared with the two-stage scheme. As shown in the figure, when $p \leq 0.4$, the mean detection time of the proposed multi-resolution scheme with bandpass sampling is smaller than one-stage schemes. As p increases, the mean detection time for one-stage schemes becomes comparable to the two-stage schemes, as the key advantage of reduced switching times for the latter is lost.

Also, the mean detection time for the proposed scheme only decreases slowly as p increases. Because $\gamma_{opt} = 3$ when $p \leq 0.45$, there are about 3 channels in each block. The time required for the first stage is independent of p according to (2.22). And $T_i + T_{sw} \leq \overline{T}_{IB} \leq 3(T_i + T_{sw})$. So the mean detection time is insensitive in the range $p < 0.5$.

The impact of SNR on the mean detection time is illustrated in Fig.2.4. $p = 0.1$. As the SNR decreases towards 5 dB, the mean detection time begins to increase rapidly. Based on this, the multi-resolution sensing scheme with energy detector is seen to sustain noise up to 5 dB. The constraints $P_d \geq 0.9$ and $P_{fa} \leq 0.1$ cannot be satisfied when SNR$\leq$ 5 dB, due to the noise aggregation in the block.

2.4.2 Optimal Points for Different SNR and Occupancy

The relationship between the optimal γ and p is discussed in this section. SNR= 10 dB. As p increases, γ_{opt} increases. This is because as p increases, $\overline{T}_{stageII} \approx T_i + T_{sw}$, and $\overline{S} \approx 1$, the mean detection time $\overline{T}_d$ is proportional to $\frac{N}{\gamma}$. Intuitively, as there are more idle channels in the spectrum when p is large, the overall detection time is dominated by the first (coarse resolution) stage instead of the second stage. Hence minimizing the detection time is equivalent to optimizing detection time of the first stage, i.e., minimize $\frac{N}{\gamma}$ subject to the constraints, according to (2.22). The optimal sub-sampling factor $\gamma = 3$ when p is low. Therefore, the mean detection time is minimized when the bandpass sampling module samples at 3x sub-Nyquist rate. For different SNR range in Fig.2.4, γ_{opt} remains the same. On the other hand, the SNR has more impact on δ_{opt} than γ_{opt}.

Fig. 2.5 illustrates the optimal δ with respect to p. As p increases, δ_{opt} decreases monotonically. Since more channels are idle, the average power in a block is lower. To satisfy the constraints of P_d and P_{fa}, the threshold δ in energy detection is thus also lower. The sudden drop when p increases from 0.4 to 0.5 corresponds to the increased γ as discussed above. As shown in Fig. 2.6, the optimal δ increases linearly when SNR increases for a fixed σ_n, because the average power in a block also increases linearly with SNR.

2.4.3 Receiver Operating Characteristics

To have a better sense of the advantage of joint optimization instead of single parameter optimization, we change the value of the sub-sampling factor γ and take various value of the energy detection threshold δ to explore the minimized mean detection time. Performance of two different approach in the scenario where SNR= 10 dB and $p = 0.2$ is shown in Fig.2.7. The lower bound of mean detection time in (2.27) with $\gamma = 3$ is also depicted. The line with 'x' gives the mean detection time for $\gamma_{opt} = 3$, which is close to the theoretical lower bound. The circle line shows the performance for $\gamma = 4$. It can be observed the optimal δ is different for the two γs, and the optimized mean detection time for the joint optimal point γ_{opt} is lower than for $\gamma = 4$.

The overall receiver operating characteristics for the multi-resolution sensing scheme

with bandpass sampling is shown in Fig.2.8 for $\gamma = 3$ and $\gamma = 4$ with SNR= 10 dB and $p = 0.3$. The results indicate that the scheme with $\gamma_{opt} = 3$ has better performance than $\gamma = 4$. With the same P_{fa}, $\gamma_{opt} = 3$ yields higher P_d than $\gamma = 4$ due to smaller γ resulting in higher block SNR.

Since the sampling rate is related to the power consumption of the cognitive radio unit, there is a trade-off between the power consumption and the performance of the multi-resolution scheme (i.e. the mean detection time and the $P_d - P_{fa}$ performance). For higher γ, the receiver ADC samples at a lower rate, and the power consumption is lower. Nevertheless, as shown in Fig.2.7 and Fig.2.8, the mean detection time is also higher for a higher γ, and the $P_d - P_{fa}$ performance is worse.

2.4.4 Validation with Real World Spectrum Data

We conducted a performance evaluation of our multi-resolution scheme using the (publicly available) measurements of the power spectrum taken in a modern office building in Aachen, Germany, during 5.5 days in 2007 [84]. The spectral band observed is 20MHz-1.64GHz (center frequency 770MHz). The RF signal was amplified, down converted to the base band, and processed by a Agilent E4440A spectrum analyzer. The analyzer produces a 8192-point FFT on the signal, where each FFT bin represents a 200 kHz channel. The magnitude squared FFT value represents the instantaneous spectrogram (power spectrum) over a channel. A full scan of the 8192 channels takes about 1.8 second. We use the data measured in 6 time slots (12:37am - 12:40am, 2:37am - 2:40am, 4:37am - 4:40am, 6:37am - 6:40am, 8:37am - 8:40am, 10:37am - 10:40am) during the first 5 days to compare the performance of the three schemes.

The following assumptions are used in data calibration:

i. The channel status is constant for a 30 second period;

ii. The channel status is the same for the same time of the day over the 5 day observation period;

iii. Adjacent channel status is coherent (fully correlated), i.e. if both of the adjacent channels to a reference channel are occupied by a primary user, it is assumed this channel

is also occupied.

Since the true status of licensed users in this band is not available, the "ground truth" for the primary users is estimated by the following steps:

Step 1: Pick a 27-second period (15 snapshots $Sn_1, Sn_2, ..., Sn_{15}$), and conduct energy detection with threshold -111 dBm on each snapshot to get 15 estimates of status on the j^{th} day $\hat{O}_{1,j}, \hat{O}_{2,j}, ..., \hat{O}_{15,j}$,

$$\hat{O}_{i,j}(k) = \{ \begin{array}{ll} 1 & if\ P_{i,j}(k) \leq -111 \\ -1 & if\ P_{i,j}(k) > -111 \end{array} \quad k = 1, ..., 8192 \tag{2.28}$$

where $P_{i,j}(k)$ is the measured power in the k^{th} channel of Sn_i on the j^{th} day.

Step 2: Use "boosting" [35] to estimate the status of the 8th snapshot Sn_8. Let the estimation of Sn_8 on the j^{th} day be $\hat{O}_d(j) = sign(\sum_{i=1}^{15} w_i \hat{O}_{i,j})$, where $w_i = 8 - |i - 8|$, i.e. the nearest time slots are given more weights than the time slots far from Sn_8.

Step 3: Exploit the correlation of the status in the same time slots of different days $\hat{O}_d = sign(\sum_{j=1}^{5} \hat{O}_d(j))$.

Step 4: If $sign(\hat{O}_d(k-1)) = sign(\hat{O}_d(k+1)) = -1$, i.e. both the two adjacent channels of a chosen channel are detected as occupied in $\hat{O}_d$, $\hat{O}_{gt}(k) = -1$. This gives the final estimated "real" status $\hat{O}_{gt}$.

The three schemes - one-stage serial search, two-stage, and multi-resolution with band-pass sampling - are applied to the Aachen dataset in these six time windows. The detection result is compared with $\hat{O}_{gt}$ to obtain P_d and P_{fa} and averaged over the six windows. Energy detection is used with threshold individually optimized to minimize detection time for all the three schemes. Because $p \approx 0.5$, $\alpha = 4$ and $\gamma = 4$ are used in the two-stage, and the multi-resolution scheme, respectively. The average P_d, P_{fa}, and the corresponding mean detection time to find an idle channel for the three schemes are shown in Table. 2.3. The multi-resolution scheme results in *significantly* lower detection time (note that the first 2500 channels are all occupied in the data). P_d and P_{fa} are approx. the same for serial search and the multi-resolution schemes (because for all the blocks $X = 1$ and hence $P_d \approx 1$ and $P_{fa} \approx 0$ in Stage I for the proposed scheme), enabling a fair comparison.

	P_d	P_{fa}	T_d (ms)
One-stage serial search	0.91	0.08	59.5
Two-stage with $\alpha = 4$	0.89	0.11	14.7
Multi-resolution with $\gamma = 4$	0.91	0.08	2.2

Table 2.3: Comparison of the three schemes using real world data

2.5 Conclusion

A multi-resolution sensing scheme with bandpass sampling is proposed to detect idle channels in cognitive radio. Bandpass sampling technique enables sub-Nyquist ADC and simplifies the receiver circuits design. To minimize the mean detection time under the constraints of P_d and P_{fa}, the sub-Nyquist sampling factor and the threshold in energy detection are optimized jointly. Numerical results show that the proposed algorithm achieves lower mean detection time compared to one stage algorithms when channel occupancy rate is low. And it is a promising alternative when Nyquist rate ADC is not feasible.

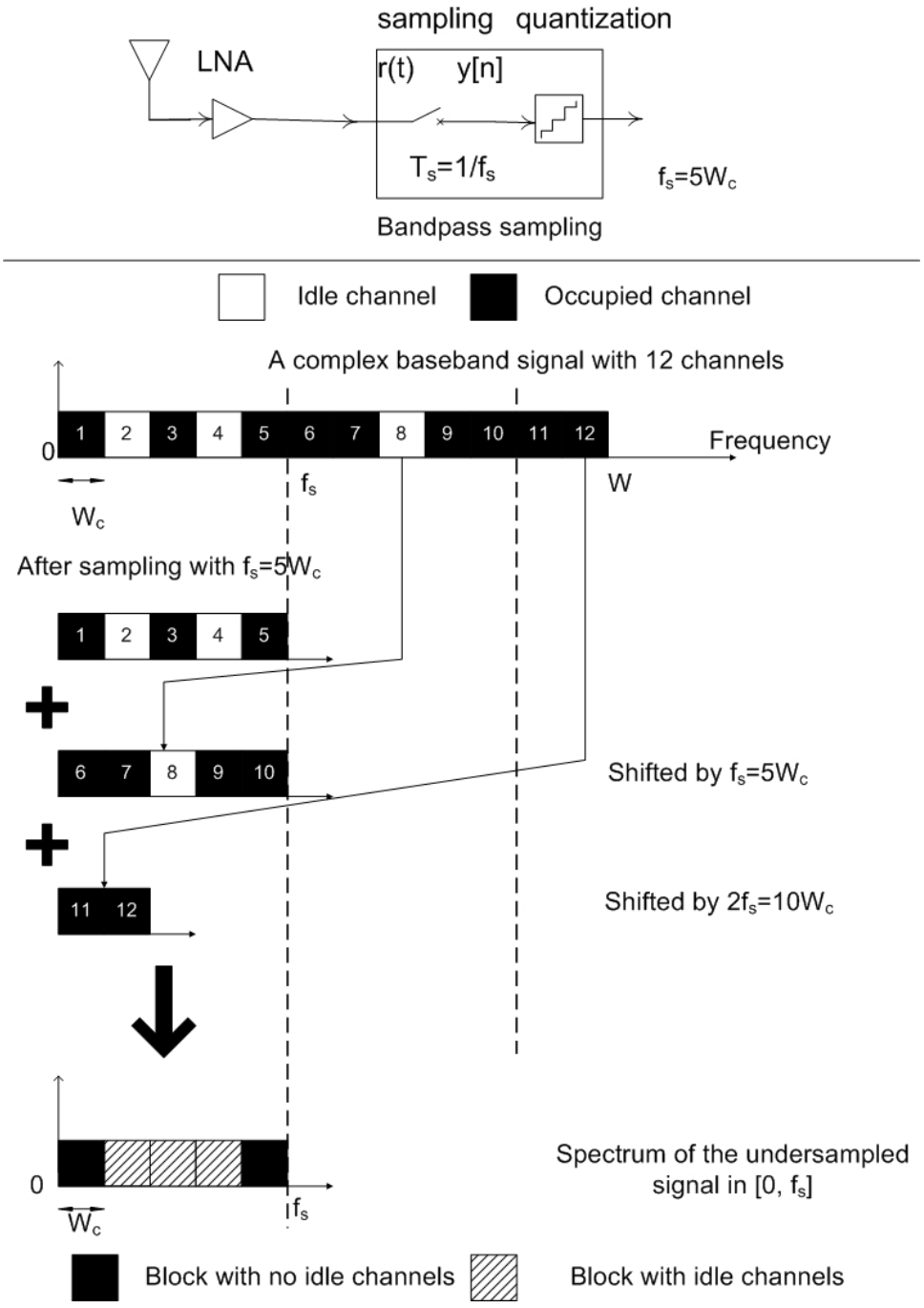


Figure 2.1: Sub-Nyquist rate sampling: a signal with 12 channels sampled at $5W_c$

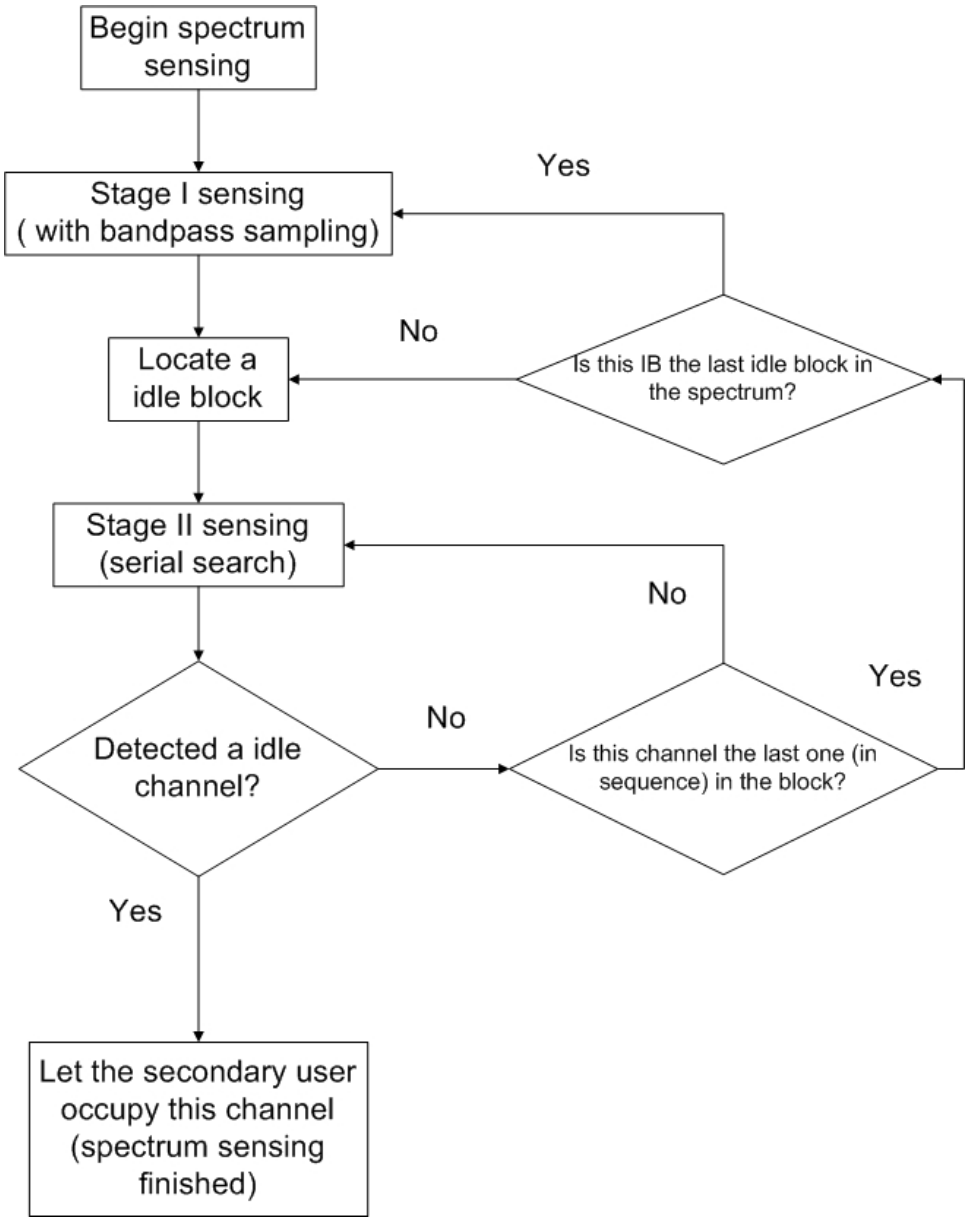


Figure 2.2: Flow chart of the multi-resolution sensing scheme

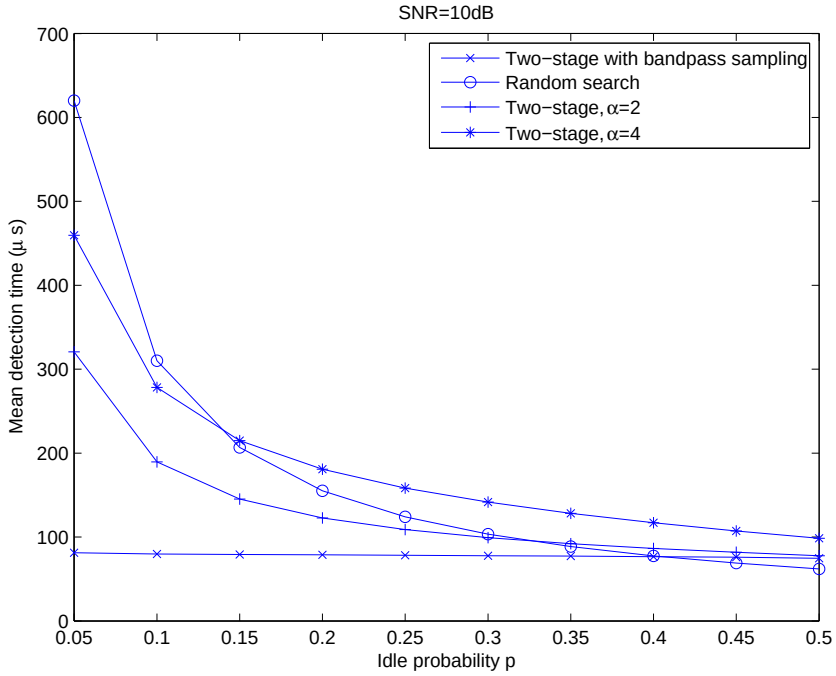


Figure 2.3: Mean detection time for the one and two stage sensing schemes versus p

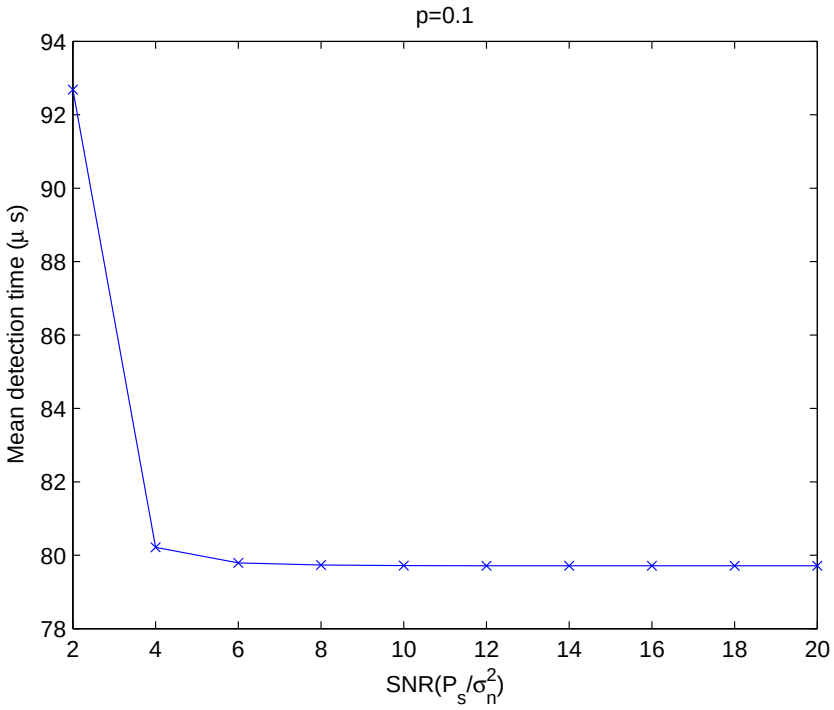


Figure 2.4: Mean detection time versus SNR

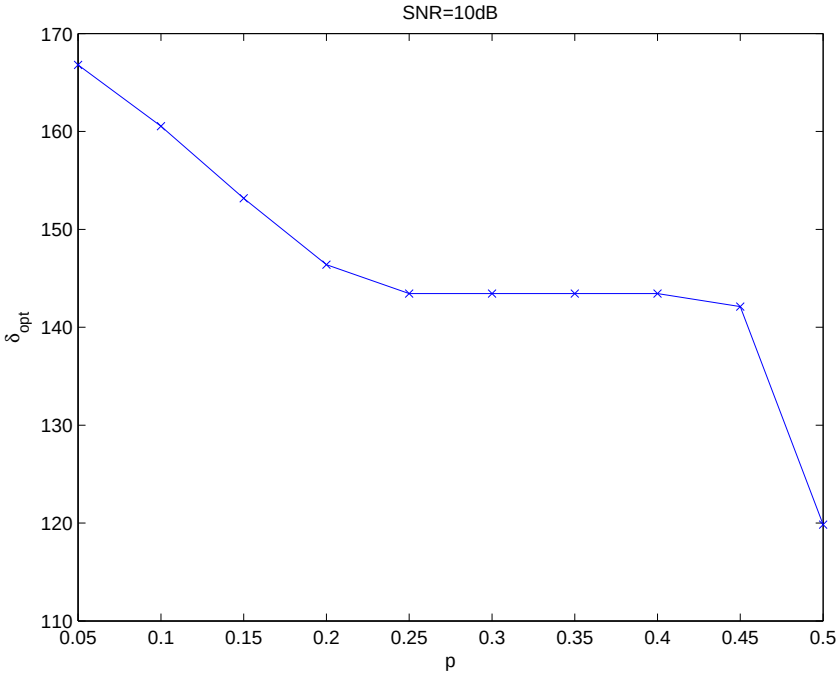


Figure 2.5: Optimal δ versus p

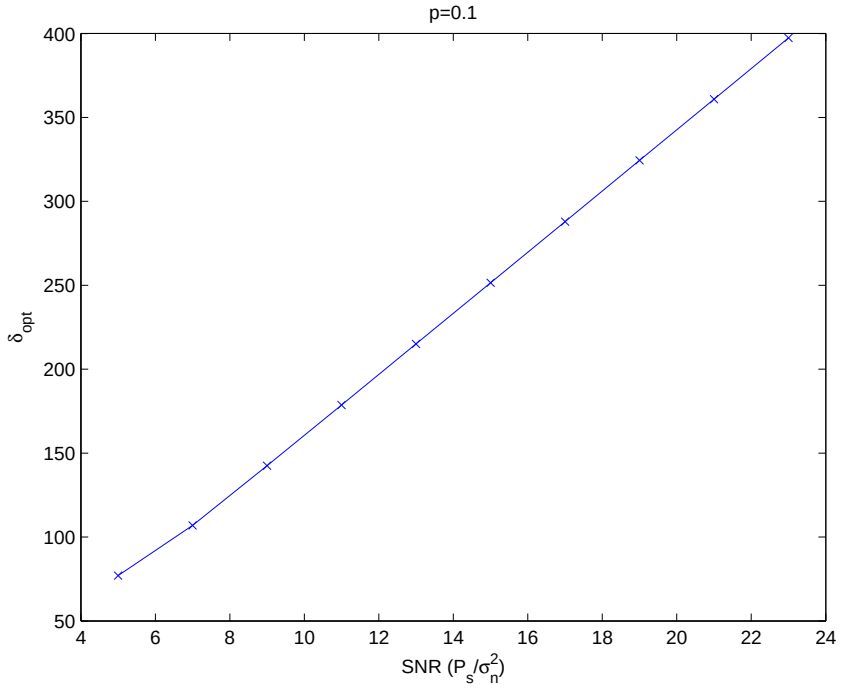


Figure 2.6: Optimal δ versus SNR

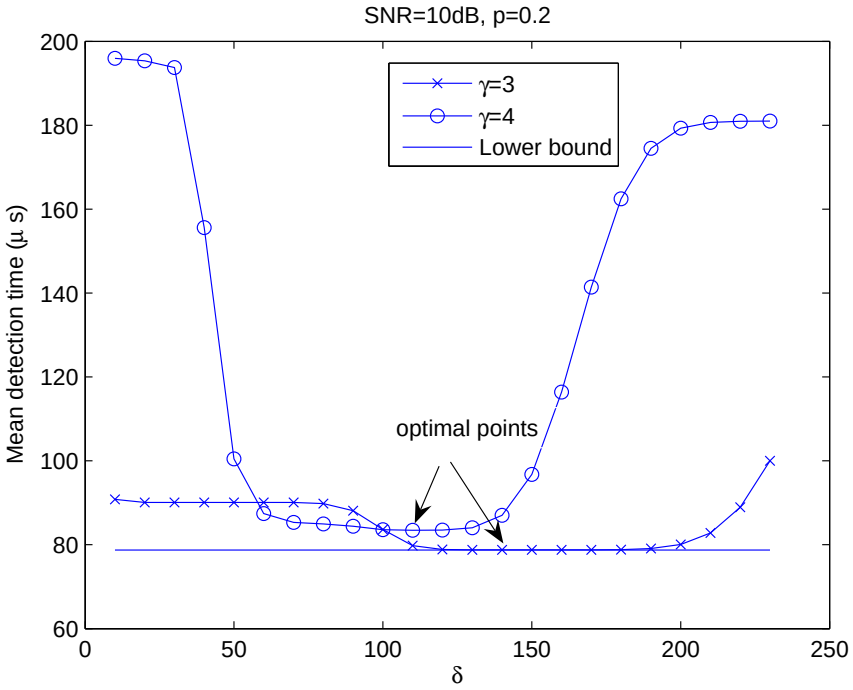


Figure 2.7: Optimal operation points

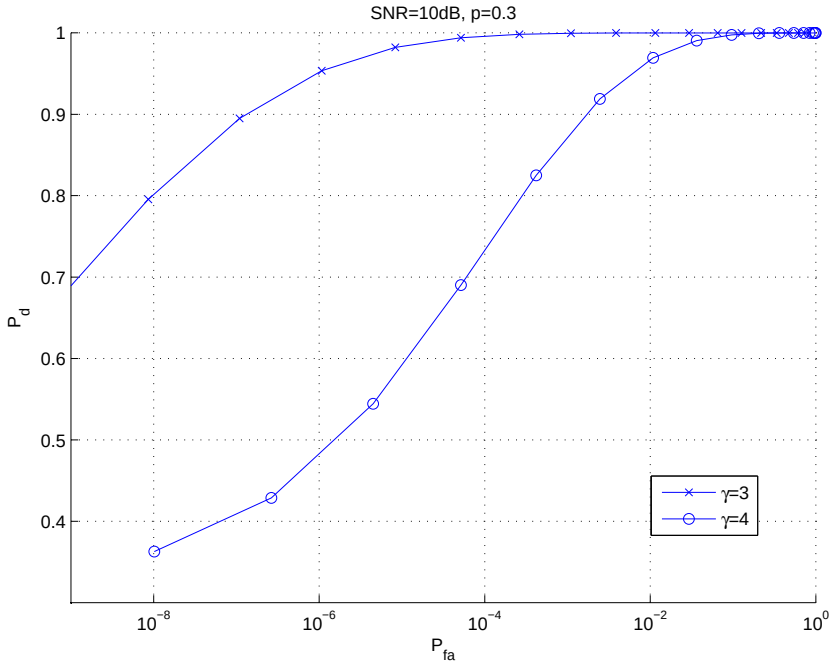


Figure 2.8: Receiver operating characteristics of the multi-resolution sensing scheme

Chapter 3

COMPRESSIVE SPECTRUM SENSING WITH BANDPASS SAMPLING

3.1 Introduction

In this chapter, we seek to detect all the idle channels in the spectrum, given most of the channels are free. We unveil a *new, compressive* architecture based on sub-Nyquist [1] sampling ADCs. Our proposal thus treads a middle ground by combining the strengths of the two architectures, in terms of achieving fast detection, while requiring a lower rate ADC.

Of late, applications of CS to spectrum estimation in cognitive radio has appeared in [75], [53], [51], and [27]. Most of the previous works focus on algorithm development for reconstruction, such as [75] and [51], while in this paper, we emphasize the architectural aspects of compressive detectors and its impact on performance. Therefore, for baseline, we compared our architecture performance with the preferred architectures in prior art. Various receiver architectures for spectrum sensing with sub-Nyquist sampling have been proposed based on compressive sensing theory. These architectures can be divided into two classes: non-uniform sampler and random pre-integrator [13]. Both receiver architectures consist of sub-Nyquist sampling ADCs to obtain an observation matrix $\mathbf{\Phi}$ that satisfies RIP. Nevertheless, these sub-Nyquist sampling approaches based on CS can provide accurate results only if *most of the channels in the spectrum are idle*, i.e., $\bar{p} \approx 1$, where $\bar{p}$ denotes the average probability of a channel being idle. If the number of occupied channels is close to the number of idle channels, i.e., $\bar{p} \approx 0.5$, current CS-based reconstruction approaches fail to provide accurate channel status [13]. In addition, compressive detection incurs high computational complexity to solve (1.2), leading to extended signal processing time. Furthermore,

[1] The sampling is sub-Nyquist relative to the rate required to sample the (bandpass) spectra treated as a low pass signal.

although the sampler in the ADC operates at a sub-Nyquist rate, the detector actually requires Nyquist rate *analog components* as discussed in Sec. 3.1.1 [54]. For the non-uniform sampler, a Nyquist rate clock is required for the shifters; for the random pre-integrator, a Nyquist rate random generator is required.

3.1.1 Non-uniform Sampling and Random Pre-integration: Compressive Detection Approaches

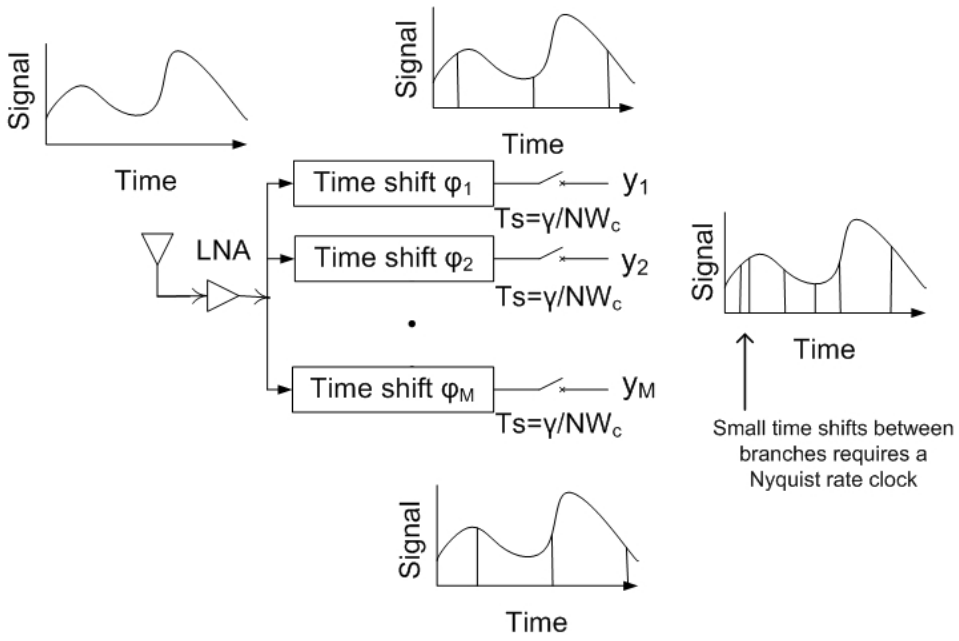


Figure 3.1: Architecture of a non-uniform sampler

The architecture of a non-uniform sampler is shown in Fig.3.1 [44][52] where the wideband input signal is sampled at a sub-Nyquist rate $f_s = \frac{1}{T_s} = \frac{NW_c}{\gamma}$ by M branches, each with a different time shift misalignment ϕ_i, $i = 1, ..., M$, where $\gamma > 1$ is the sub-sampling factor. In terms of analog implementation, the non-uniform sampler operates at a sub-Nyquist rate, but requires a Nyquist rate clock to synchronize the shifters on each branch. On the other hand, an example of a random pre-integrator architecture is shown in Fig. 3.2 [54]. In a random pre-integrator [39][77], a demodulator multiplies the received (analog)

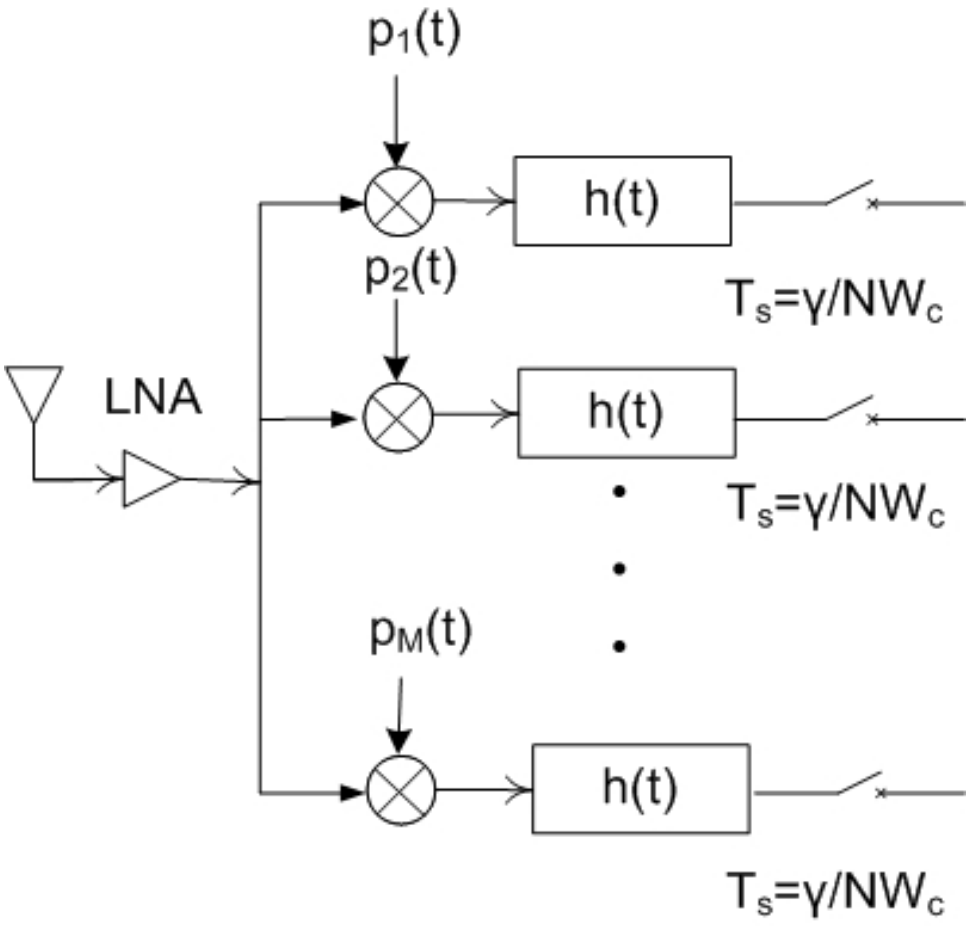


Figure 3.2: Architecture of a random pre-integrator

signal with a random chipping sequence; the resulting product is integrated over a symbol duration and sampled at a sub-Nyquist rate on each of the M branches. The chipping sequence generator is required to work at greater or equal to the Nyquist rate to produce a $\mathbf{\Phi}$ with i.i.d. Bernoulli or Gaussian entries.

Denote the signal processing time for (1.2) as T_{l1}, then the mean detection time for the non-uniform sampler is

$$T_{d,CS} = (N \cdot N_i)/(N \cdot W_c) + N_i T_{l1} = N_i/W_c + N_i T_{l1} \tag{3.1}$$

where T_{l1} is determined by the computational complexity of the signal processing algo-

rithms. Using the non-uniform sampling architecture, the complexity for solving (1.2) is $O(M^2N^{1.5})$ [89]. In [46], as $M \sim O((N-\bar{p}N)log(1/(1-\bar{p})))$, $T_{l1} \sim O(((1-\bar{p})N)^3)$. When N_i spectrum snapshots are reconstructed and averaged to generate the power spectrum, the digital signal processing time is $O(((1-\bar{p})N)^3 N_i)$ for both the random pre-integrator and non-uniform sampling architectures. Note that (1.2) requires the sensing matrix $\mathbf{\Phi\Psi}$ to be stored on-chip, yielding a memory requirement of $O(MN)$. However, the storage required for the non-uniform sampler is $O(M)$. Since $\mathbf{\Phi\Psi}$ is a subset of the IDFT matrix, only the indices for the M chosen rows need to be stored, and $\mathbf{\Phi\Psi}$ can be reconstructed using only the indices.

3.1.2 Contributions

Our work is a synthesis of the notion of Compressive Sensing within a bandpass sampling architecture. The traditional bandpass sampling receiver chain is shown in Fig.3.3, where the *complex baseband* signal $r(t)$ is sampled at a sub-Nyquist rate $f_s = \frac{1}{T_s} = \frac{NW_c}{\gamma}$ to obtain $y[n] = r(nT_s)$, where $\gamma > 1$ is the sub-sampling factor [81]. Note that for compactness, only the *analytic* signal that occupies the positive frequency spectrum, is shown. This architecture achieves frequency down-conversion by undersampling $r(t)$ (viewed as a wideband lowpass signal), where the sampling rate is determined by the bandwidth of the desired bandpass signal (i.e., conforming to the bandpass Nyquist sampling theorem) instead of the carrier frequency [3]. Owing to the reduction of the sampling rate, bandpass sampling simplifies the design of the local oscillator (LO), leading to better performance than a mixer for down-conversion [67].

In Sec.4.2, a spectrum sensing scheme with a bandpass sampling architecture is proposed. Here, the entire spectrum (in principle) can be sensed *simultaneously*, reducing the need for LO scanning. The proposed architecture switches between different (sub-Nyquist) sampling rates to obtain an observation matrix that satisfies RIP; no Nyquist-rate analog component is needed. Instead of reconstructing each snapshot of the amplitude spectra, the power spectrum is reconstructed directly by l_1 minimization, mitigating the delay of intense signal processing.

In Sec.3.3, the performance of the proposed architecture is analyzed. The mean detection time, the computational complexity, and the required memory for storing the observation matrix are compared with the non-uniform sampler and the random pre-integrator. In addition, the detection rate and the false alarm rate for the proposed scheme is analyzed. A low detection rate will cause interference to primary users, while a low false alarm rate is necessary for a high spectral utilization rate. Numerical results for the mean detection time are presented as a function of the number of channels to be sensed. Impact of Signal-to-Noise Ratio (SNR) and $\bar{p}$ is explored. Finally, in Sec.3.4, the proposed scheme is validated using real-world spectrum measurements obtained in Aachen, Germany.

3.2 System Description

3.2.1 System Model

Assume that the band of interest consists of N equi-spaced channels; the central frequency of the i^{th} channel is $f_i = i\,W_c - \frac{W_c}{2}$, where W_c is the bandwidth for each channel. A binary variable, O_i is used to denote the status of the i^{th} channel, where $O_i = 0(1)$ means the channel is busy (idle). We invoke an independent Bernoulli model for channel status, that is supported by real-world data as described in [31], i.e.,

$$Pr(O_i = 1) = p_i, i = 1, ..., N \tag{3.2}$$

Thus, the average signal power $\bar{P}_{s,i}$ in the i^{th} channel is described as a Bernoulli variable, with $Pr(\bar{P}_{s,i} = 0) = p_i$ and $Pr(\bar{P}_{s,i} = \bar{P}_{r,i}) = 1 - p_i$, where p_i is the probability of the i^{th} channel being idle. $\bar{P}_{r,i}$ is the signal power in the i^{th} channel at the receiver when the channel is occupied by the primary user. The *average* idle channel probability over the entire band is given by $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$.

3.2.2 Compressive sensing architecture

Following the discussions in Chapter 2, each element in $\mathbf{Y}$ is the average of the amplitude spectra over a block of γ channels in $\mathbf{R}$. Let $\mathbf{R} = \mathbf{S} + \mathbf{V}$, where $\mathbf{S} = [S_1, \ldots, S_N]^T$ is the signal component and $\mathbf{V} = [V_1, \ldots, V_N]^T$ is the noise component, $E(|S_i|^2) = \bar{P}_{s,i}$, and

$E(|V_i|^2) = \bar{P}_{v,i}$, where $\bar{P}_{s,i}$ ($\bar{P}_{v,i}$) is the average signal (noise) power in the i^{th} channel, respectively. Then (2.5) can be written as

$$\mathbf{Y} = \frac{1}{\gamma}\mathbf{\Phi}(\mathbf{S} + \mathbf{V}) \tag{3.3}$$

The sequences S_i, V_i are assumed to consist of independent variables (corresponding to different channels) that are mutually independent. It is natural that $E(V_i) = 0$ and for symmetric signal constellations,

$$E(S_i) = 0 \tag{3.4}$$

as well. Then R_i and R_j are independent, $\forall\, i \neq j$, and $E(R_i) = E(S_i) + E(V_i) = 0$, $E(|R_i|^2) = \bar{P}_{s,i} + \bar{P}_{v,i}$. Typically the time samples $r[n]$ are complex Gaussian variables, whose Fourier coefficients $R_i \sim CN(0, (\bar{P}_{s,i} + \bar{P}_{v,i})/2)$, $i = 1, \ldots, N$ [2]. The expectation of the power in the m^{th} block is

$$\begin{aligned}
E(Y_m Y_m^*) &= E(\frac{1}{\gamma^2}|\sum_{j'=1}^{\gamma} R_{j'}|^2) \\
&= \frac{1}{\gamma^2}E((\sum_{j'=1}^{\gamma} R_{j'})(\sum_{j'=1}^{\gamma} R_{j'}^*)) \\
&= \frac{1}{\gamma^2}E(\sum_{j'=1}^{\gamma} |R_{j'}|^2 + \sum_{j_1' \neq j_2'} R_{j_1'} R_{j_2'}^*) \\
&= \frac{1}{\gamma^2}\sum_{i'=1}^{\gamma} E(|R_{i'}|^2) + \sum_{i' \neq j'} E(R_{i'})E(R_{j'}) \\
&= \frac{1}{\gamma^2}\sum_{j'=1}^{\gamma} E(|R_{j'}|^2) \\
&= \frac{1}{\gamma^2}\sum_{j'=1}^{\gamma} \bar{P}_{s,j'} + \bar{P}_{v,j'}
\end{aligned} \tag{3.5}$$

where $\bar{P}_{s,j'}$ and $\bar{P}_{v,j'}$ are the average signal power and noise power in the j'^{th} channel in the m^{th} block, respectively.

We propose to use the bandpass sampling architecture shown in Fig.2.1 to sense the spectrum K times, each for a different sub-Nyquist sampling rate, in order to obtain a $\mathbf{\Phi}$

that satisfies the RIP. An equivalent architecture is shown in Fig.3.4 where the incoming signal is fed into K branches. The k^{th} branch consists of an ADC sampling at $\frac{NW_c}{\gamma_k}$, followed by a square-law device and an integrator. Note that instead of the multi-branch architecture, a single branch detector as in Fig.2.1 that sequentially switches between different sampling rates is also feasible. However, the resulting mean detection times are different. Clearly, the multi-branch architecture yields a reduced sensing duration due to the inherent parallelism, at the cost of chip complexity. Our analysis for mean detection time in Sec.3.3 is based on the single branch architecture.

The measurement $\mathbf{Y}_{k,n}$ for the k^{th} sampling rate and the n^{th} snapshot is used to create the spectrogram, denoted by a $\frac{N}{\gamma_k} \times 1$ vector $\mathbf{w}_{k,n}$, where $\mathbf{w}_{k,n}(i) = \gamma_k^2 |\mathbf{Y}_{k,n}(i)|^2$, $n = 1, \ldots, N_i$, are i.i.d corresponding to different snapshots. Now

$$E(\mathbf{w}_{k,n}) = \mathbf{\Phi}_k \begin{bmatrix} \bar{P}_{s,1} + \bar{P}_{v,1} \\ \vdots \\ \bar{P}_{s,N} + \bar{P}_{v,N} \end{bmatrix} \tag{3.6}$$

where $\mathbf{\Phi}_k$ is constructed according to (2.6),

$$\begin{aligned} & Var(\mathbf{w}_{k,n}(i)) \\ = \; & E((\mathbf{w}_{k,n}(i))^2) - (E(\mathbf{w}_{k,n}(i)))^2 \\ = \; & E(\gamma_k^4 |\mathbf{Y}_{k,n}(i)|^4) - (\textstyle\sum_{j'=1}^{\gamma_k} (\bar{P}_{s,j'} + \bar{P}_{v,j'}))^2 \\ = \; & E(|\textstyle\sum_{j'=1}^{\gamma_k} R_{j'}|^4) - (\textstyle\sum_{j'=1}^{\gamma_k} (\bar{P}_{s,j'} + \bar{P}_{v,j'}))^2 \\ = \; & E((\textstyle\sum_{j_1'=1}^{\gamma_k} R_{j_1'})^2 (\textstyle\sum_{j_2'=1}^{\gamma_k} R_{j_2'}^*)^2) - (\textstyle\sum_{j'=1}^{\gamma_k} (\bar{P}_{s,j'} + \bar{P}_{v,j'}))^2 \\ = \; & \textstyle\sum_{j'=1}^{\gamma_k} E(|R_{j'}|^4) \\ = \; & (\textstyle\sum_{j'=1}^{\gamma_k} (\bar{P}_{s,j'} + \bar{P}_{v,j'}))^2 \end{aligned} \tag{3.7}$$

For $i_1 \neq i_2$,

$$
\begin{aligned}
& Cov(\mathbf{w}_{k,n}(i_1), \mathbf{w}_{k,n}(i_2)) \\
= \quad & E((\mathbf{w}_{k,n}(i_1))(\mathbf{w}_{k,n}(i_2))) \\
& -(E(\mathbf{w}_{k,n}(i_1)))(E(\mathbf{w}_{k,n}(i_2))) \\
= \quad & 0
\end{aligned} \tag{3.8}
$$

The averaged power over N_i snapshots

$$
\bar{\mathbf{w}}_k = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{w}_{k,n} \tag{3.9}
$$

has a chi-square distribution with $2N_i$ degrees of freedom. When the number of samples N_i is large, we can use the Central Limit Theorem to approximate $\bar{\mathbf{w}}_k$ as Gaussian [11], i.e., $\bar{\mathbf{w}}_k \sim N(E(\mathbf{w}_{k,n}), \boldsymbol{\Sigma}_k/N_i)$, where the variance matrix

$$
\boldsymbol{\Sigma_k} = diag(Var(\mathbf{w}_{k,n}(1)), \ldots, Var(\mathbf{w}_{k,n}(\frac{N}{\gamma_k}))) \tag{3.10}
$$

Therefore, for the K sampling rates,

$$
\begin{bmatrix} \bar{\mathbf{w}}_1 \\ \vdots \\ \bar{\mathbf{w}}_K \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_K \end{bmatrix} \begin{bmatrix} \bar{P}_{s,1} \\ \vdots \\ \bar{P}_{s,N} \end{bmatrix} + \mathbf{n} \tag{3.11}
$$

Denote $\bar{\mathbf{w}}^{CS} = \begin{bmatrix} \bar{\mathbf{w}}_1 \\ \vdots \\ \bar{\mathbf{w}}_K \end{bmatrix}$, $\boldsymbol{\Phi}^{CS} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_K \end{bmatrix}$, $\mathbf{s}^{CS} = \begin{bmatrix} \bar{P}_{s,1} \\ \vdots \\ \bar{P}_{s,N} \end{bmatrix}$, leading to the compact form

$$
\bar{\mathbf{w}}^{CS} = \boldsymbol{\Phi}^{CS}\mathbf{s}^{CS} + \mathbf{n} \tag{3.12}
$$

where

$$\mathbf{\Phi}^{CS} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & ... \end{array} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & ... \\ 0 & 1 & 0 & 0 & 0 & ... \\ \vdots & & & & & \\ 1 & 0 & 0 & 0 & 0 & ... \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & ... \end{array}\right] \end{array} \begin{array}{c} n = 1, ..., N \\ (n \ mod \frac{N}{\gamma_1}) = 1 \\ (n \ mod \frac{N}{\gamma_1}) = 2 \\ \vdots \\ (n \ mod \frac{N}{\gamma_2}) = 1 \\ \vdots \\ (n \ mod \frac{N}{\gamma_K}) = 0 \end{array} \tag{3.13}$$

Define $\frac{N}{\gamma_0} = 0$, then $\mathbf{\Phi}^{CS}(i,j) = 1$, if $\exists\ 0 \leq l \leq K-1,\ s.t. \sum_{k=0}^{l} \frac{N}{\gamma_k} < i \leq \sum_{k=0}^{l+1} \frac{N}{\gamma_k}$, and $j \ mod \ \frac{N}{\gamma_{l+1}} = i - \sum_{k=0}^{l} \frac{N}{\gamma_k}$; otherwise $\mathbf{\Phi}^{CS}(i,j) = 0$.

The additive noise $\mathbf{n} \sim N(\mathbf{\Phi}^{CS} \begin{bmatrix} \bar{P}_{v,1} \\ \vdots \\ \bar{P}_{v,N} \end{bmatrix}, \mathbf{\Sigma}^{CS})$. Because $\bar{\mathbf{w}}_{k_1}$ and $\bar{\mathbf{w}}_{k_2}$ are derived from samples from different time slots, $\forall\, k_1 \neq k_2$, $Cov(\bar{\mathbf{w}}_{k_1}(i_1), \bar{\mathbf{w}}_{k_1}(i_2)) \approx 0$, $\forall\, i_1, i_2$, yielding $\mathbf{\Sigma}^{CS} \approx \frac{1}{N_i} diag(\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K)$.

Because $Pr(\bar{P}_{s,i} = 0) = p_i$, the average number of non-zero elements in $\mathbf{s}^{CS}$ is $(1 - \bar{p})N$. It is proved in [40] that if $(1-\bar{p}) \ll 1$, and $\frac{N}{\gamma_1}, \ldots, \frac{N}{\gamma_K}$ are the first $K \sim O((N - \bar{p}N) log_{N-\bar{p}N} N)$ prime numbers with $(1-\bar{p})N \leq \frac{N}{\gamma_1} < \ldots < \frac{N}{\gamma_K}$, then $\mathbf{\Phi}^{CS}$ satisfies RIP. Note that a lower bound for $\bar{p}$ is required for choosing K, since $\forall\, 0 < p_{low} < p < 1$, if $\mathbf{\Phi}^{CS}$ satisfies RIP for $(2(1-p_{low})N, \delta_{2N(1-p_{low})})$, it satisfies RIP for $(2(1-p)N, \delta_{2N(1-p_{low})})$. Detection of the channel status $\mathbf{O} = [O_1, \ldots, O_N]^T$ in (3.2) is done by first solving for the signal powers from the following convex optimization problem

$$\begin{aligned} &minimize \ \|\hat{\mathbf{s}}\|_1 \\ &subject \ to \ \|\bar{\mathbf{w}}^{CS} - \mathbf{\Phi}^{CS}\hat{\mathbf{s}}\|_2 \leq \epsilon \end{aligned} \tag{3.14}$$

where $Pr(\|\mathbf{n}\|_2 \leq \epsilon) \geq \lambda$, $\lambda \approx 1$. Then each element in $\hat{\mathbf{s}} = [\hat{P}_{s,1}, \ldots, \hat{P}_{s,N}]^T$ is compared to a power threshold β to determine if the channel is occupied, i.e., $\hat{P}_{s,i} > \beta \Rightarrow \hat{O}_i = 0$.

	Non-uniform sampler	Random pre-integrator	Bandpass sampling
Mean detection time	$N_i/W_c + N_i T_{l1}$	$N_i/W_c + N_i T_{l1}$	$K(N_i/W_c + T_{sw}) + T_{l1}$
Nyquist rate analog component	Yes	Yes	No
Storage (projection matrix)	$O((N - \bar{p}N)log(1/(1-\bar{p})))$	$O(N(N - \bar{p}N)log(1/(1-\bar{p})))$	$O((N - \bar{p}N)log_{N-\bar{p}N}N)$

Table 3.1: Comparison between CS-based architectures

3.3 Performance Analysis

In this section, the performance of the proposed compressive spectrum sensing architecture with bandpass sampling is analyzed. The mean detection time and the storage required for the observation matrix is compared with the non-uniform sampler and the random pre-integrator in Table 3.1.

3.3.1 Mean Detection Time and Computational Complexity

The mean detection time for the proposed compressive sensing architecture consists of three parts: integration time, time for switching ADC sampling rates, and latency for computation of the spectrogram. Since N_i samples are used in determining status for each folded channel in $\bar{\mathbf{w}}_k$, the total number of samples needed equals $N_i N/\gamma_k$. Hence, the total integration time for the k^{th} sampling rate is $\frac{N_i N}{\gamma_k}\frac{\gamma_k}{NW_c} = N_i/W_c$ and for the entire sensing process is KN_i/W_c. Denote the switching time between two different sampling rates as T_{sw}. The spectrogram $\bar{\mathbf{w}}_k$ for each sampling rate can be generated during the analog circuit switching time, because the FFT size for each sampling rate is $\frac{N}{\gamma_k} \ll N,\ k = 1, \ldots, K$, and the signal processing time for FFT and averaging is negligible compared to T_{sw}. Therefore, the major component of the latency for computation of $\hat{\mathbf{s}}$ is the signal processing time T_{l1} required by (3.14). Thus the mean detection time for the proposed sensing scheme is

$$T_{d,bps} = K(N_i/W_c + T_{sw}) + T_{l1} \tag{3.15}$$

The number of rows M of $\mathbf{\Phi}^{CS}$ in (3.13) is KC_L, where C_L is the average of the first K prime numbers no smaller than $N(1-\bar{p})$. In [40], it is proved that $C_L \sim O((N(1-\bar{p}))(log_{N-\bar{p}N}N)ln((N-\bar{p}N)log_{N-\bar{p}N}N))$. As $K \sim O((N-\bar{p}N)log_{N-\bar{p}N}N)$, $M \sim O(C_L(N-\bar{p}N)log_{N-\bar{p}N}N)$. l_1 minimization is applied to compute the averaged power spectrum directly, instead of the individual amplitude spectra for each snapshot. Thus the computational complexity associated with (3.14) is $O(M^2N^{1.5})$, independent of N_i, leading to shorter signal processing time compared to the non-uniform sampler and the random pre-integrator.

Fig.3.5 illustrates the mean detection time for serial search with the heterodyne architecture, non-uniform sampler, random pre-integrator, and the proposed bandpass sampling architecture with $W_c = 200$ kHz, $T_{sw} = 50\mu s$, and $N_i = 1000$. As shown in [46], $T_{l1} \approx ((1-\bar{p})N)^3\mu s$. $\bar{p}$ is chosen to be 0.9 so that the spectrum satisfies the sparsity requirement of CS. The results show that the bandpass sampling architecture consistently outperforms the serial search scheme for heterodyne architecture, by reducing analog circuit switching steps. For the bandpass sampling architecture, as N increases, an increased number of sampling rates is required, leading to the increased mean detection time. For the non-uniform sampler and the random pre-integrator, as N increases, the digital signal processing time becomes the major component in mean detection time, where l_1 minimization is applied to reconstruct the amplitude spectra and is executed N_i times. When the signal processing duration is longer than the circuit switching time in the bandpass sampling architecture, the mean detection time of the non-uniform sampler and the random pre-integrator exceeds that of the bandpass sampling architecture, because $K \ll N$.

In terms of storage requirements, only the sub-sampling factors γ_k, $k = 1, \ldots, K$ need to be stored in memory, and the matrix $\mathbf{\Phi}^{CS}$ can be generated on-line according to (3.13). The storage required is smaller than the non-uniform sampler and the random pre-integrator.

3.3.2 Detection performance and receiver operating characteristics

The detection performance is influenced by the average SNR of all the channels. Assume a known signal is used for calibration, where the average SNR of the occupied channels is used for performance analysis. Denote the support of $\mathbf{s}^{CS}$ as $\Lambda = \{i : \bar{P}_{s,i} > 0\}$. The input SNR is given by

$$ISNR = \frac{\sqrt{(\sum_{i\in\Lambda} \bar{P}_{s,i})^2}}{\sqrt{(\sum_{i\in\Lambda} \bar{P}_{v,i})^2}} = C_0 \frac{\|\mathbf{s}^{CS}\|_2}{\epsilon} \tag{3.16}$$

where C_0 is determined by N and $\bar{p}$. The post-reconstruction SNR is given by

$$RSNR = \frac{\|\mathbf{s}^{CS}\|_2}{\|\mathbf{s}^{CS} - \hat{\mathbf{s}}\|_2} \tag{3.17}$$

where $\hat{\mathbf{s}}$ is output from (3.14) and $\mathbf{s}^{CS}$ consists of the actual signal powers. It is proved in [12] that with probability $\lambda \sim 1$, $\|\hat{\mathbf{s}} - \mathbf{s}^{CS}\|_2 \leq C_1\epsilon$, where $C_1 = \frac{2\alpha}{1-\rho}$, $\rho = \frac{\sqrt{2}\delta_{2\bar{L}}}{1-\delta_{2\bar{L}}}$, $\alpha = \frac{2\sqrt{1+\delta_{2\bar{L}}}}{\sqrt{1-\delta_{2\bar{L}}}}$, and $\bar{L} = (1-\bar{p})N$ [12]. Denote $C = C_0C_1$. Therefore,

$$ISNR \leq C \times RSNR \tag{3.18}$$

The post-reconstruction SNR is C times worse than the input SNR. Denote P_d and P_{fa} as detection rate and false alarm rate, respectively. Since P_d-P_{fa} performance is proportional to RSNR and RSNR is bounded at least by a function of $\delta_{2\bar{L}}$ and ISNR in (3.18), the detection rate P_d and the false alarm rate P_{fa} are determined by $\delta_{2\bar{L}}$ and ISNR, where $\delta_{2\bar{L}} \sim O(\frac{2N(1-\bar{p})}{K} log_{2N(1-\bar{p})}N)$ according to [40].

To gain additional insight into the impact of $\delta_{2\bar{L}}$ on the detection performance, the overall receiver operating characteristics for the bandpass sampling architecture are shown in Fig.3.6(a) and Fig.3.6(b), with $\bar{p} = 0.95$, $N = 100$, $N_i = 1000$, and $p_i = \bar{p}$, $i = 1, \ldots, N$. For $K = 3$, the sampling rates used are $5W_c$, $7W_c$, $11W_c$, and for $K = 4$, an additional sampling rate $13W_c$ is used, where all the sampling rates are smaller than the Nyquist rate $100W_c$. $\bar{P}_{s,i}$ are set to 10 and $\bar{P}_{v,i} = 1$, $i = 1, \ldots, N$, so ISNR = 10dB. The numerical results of the estimated detection rate $\hat{P}_d$ and the estimated false alarm rate $\hat{P}_{fa}$ are averaged over 500 iterations. In each iteration, the locations of the occupied channels are different.

As shown in Fig.3.6(a) and Fig.3.6(b), the performance for 4 sampling rates is better than 3 sampling rates, because $\delta_{2N(1-\bar{p})}$ for $K = 4$ is smaller than for $K = 3$, yielding smaller C in (3.18).

The impact of the ISNR $= \bar{P}_{s,i}/\bar{P}_{v,i}$ on the P_d-P_{fa} performance is shown in Fig.3.7. As the ISNR decreases towards 9 dB, $\hat{P}_d$ with 4 sampling rates ($5W_c$, $7W_c$, $11W_c$, and $13W_c$) begins to decrease rapidly, because the spectrum of $\mathbf{R}$ is no longer sparse for $\mathbf{\Phi}^{CS}$ with the chosen sampling rates. In contrast, all the occupied channels are correctly detected for the 5 sampling rates ($5W_c$, $7W_c$, $11W_c$, $13W_c$, and $17W_c$) over the 500 iterations for the SNR range shown in Fig. 3.7. Therefore, for a given ISNR, the detection performance of the proposed architecture can be improved by increasing the number of sampling rates, which imposes a trade-off between the detection performance and the detection time. The impact of the sparsity level of the spectrum is shown in Fig.3.8. For a smaller $\bar{p}$, an increased number of sampling rates is required for $\mathbf{\Phi}^{CS}$ to satisfy the RIP requirement as discussed in Sec.4.2, leading to a longer detection time.

3.4 Experimental Results

We conducted a performance evaluation of our bandpass sampling scheme using the (publicly available) measurements of the power spectrum taken in a modern office building in Aachen, Germany, in 2007 [84]. The spectral band observed is 20MHz-1.64GHz (center frequency 770 MHz). The RF signal was amplified, down-converted to the base band, and processed by an Agilent E4440A spectrum analyzer. The analyzer produces a 8192-point FFT on the signal, where each FFT bin represents a 200 kHz resolution width. The magnitude squared FFT value represents the instantaneous spectrogram (power spectrum) over a channel. A full scan of the 8192 channels requires about 1.8 second. The average noise level of the receiver chain is -169 dBm at 1 Hz resolution bandwidth, which is -116 dBm for the 200kHz channel. The band plan for spectrum usage in Germany was not available, so the channel status is first estimated by thresholding the original data at -111 dBm, corresponding to constant false alarm rate of 10^{-5}. The data from the final 100 channels at $12:37am$, $2:37am$, $4:37am$, $6:37am$, and $8:37am$ on Jan.27th, is used as input for the bandpass sampling architecture.

Among the 100 channels, two channels are occupied as shown in Fig.3.9(a) and Fig.3.9(b). For the bandpass sampling architecture, the 4 sampling rates used are $5W_c$, $7W_c$, $11W_c$, and $13W_c$, and for 5 sampling rates, $17W_c$ is added. To enhance signal sparsity, the average noise level of -116 dBm is removed from the original data, and then added back in the reconstructed power spectrum for comparison with the original spectrum. The proposed architecture with $K = 4$ achieves a $\hat{P}_d = 1$ while $\hat{P}_{fa} = 0.01$ for all the 5 time slots, with the same threshold of -111 dBm. With $K = 4$, the reconstructed power spectrum for $12:37am$ and $8:37am$ is shown in Fig.3.9(a) and Fig.3.9(b). Although the SNR of the reconstructed signal is worse than the original signal, both of the 2 occupied channels can be accurately identified, but a false alarm occurs in Fig.3.9(b). As shown in Fig. 3.10(a), with $K = 5$, no false alarm occurs for $8:37am$, because the reconstructed SNR is improved according to Section 3.3.

3.5 Conclusion

A compressive spectrum sensing architecture with bandpass sampling is proposed to detect idle channels in cognitive radio. The bandpass sampling technique enables sub-Nyquist ADCs and simplifies the receiver circuits design. Compared to other compressive sensing architecture, the proposed architecture does not have any Nyquist rate analog circuit components. Numerical results show that the proposed algorithm achieves lower mean detection time compared to the serial search scheme when the channel occupancy rate is low and is a promising alternative when Nyquist rate ADC is not feasible.

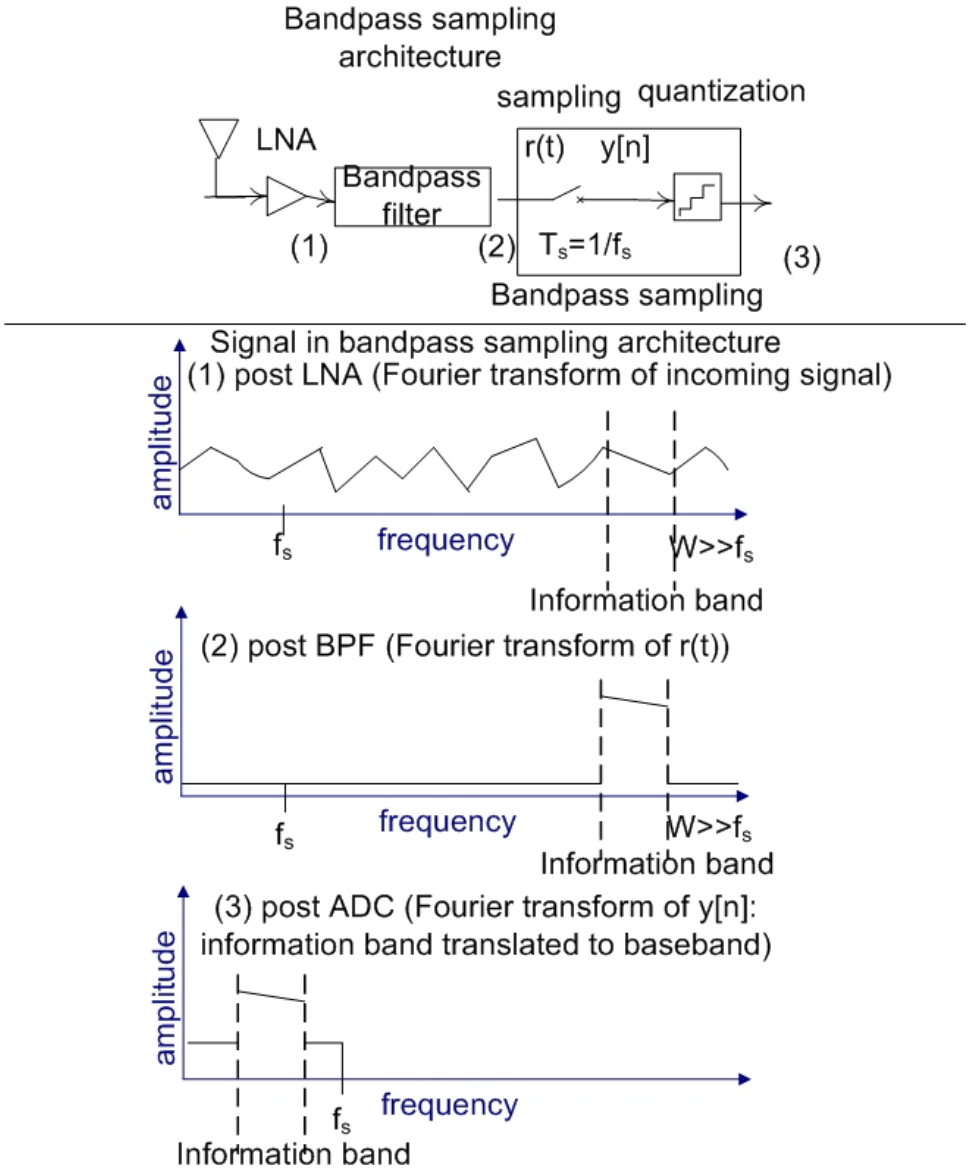


Figure 3.3: Block diagram of a bandpass sampling architecture for direct down-conversion

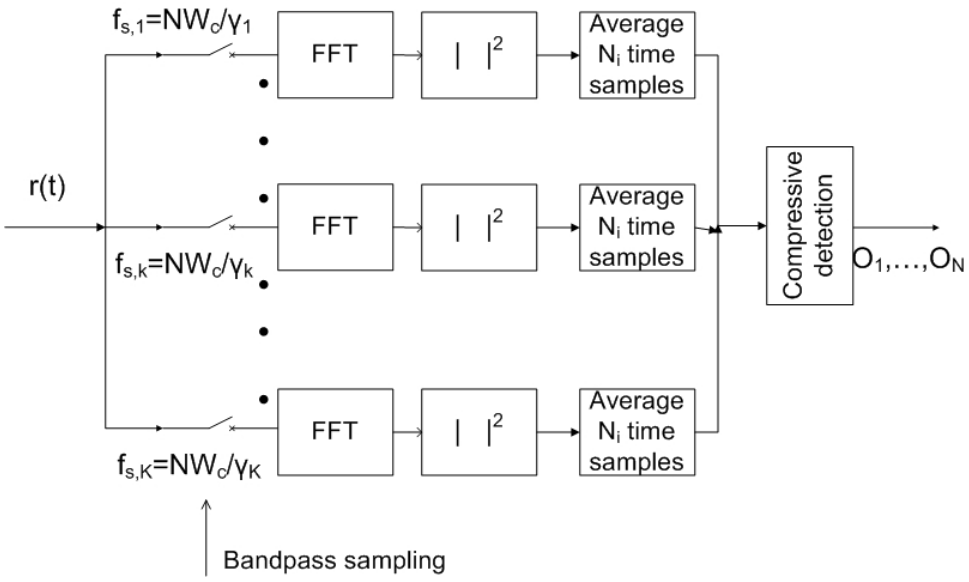


Figure 3.4: Block diagram of the compressive detection scheme with bandpass sampling

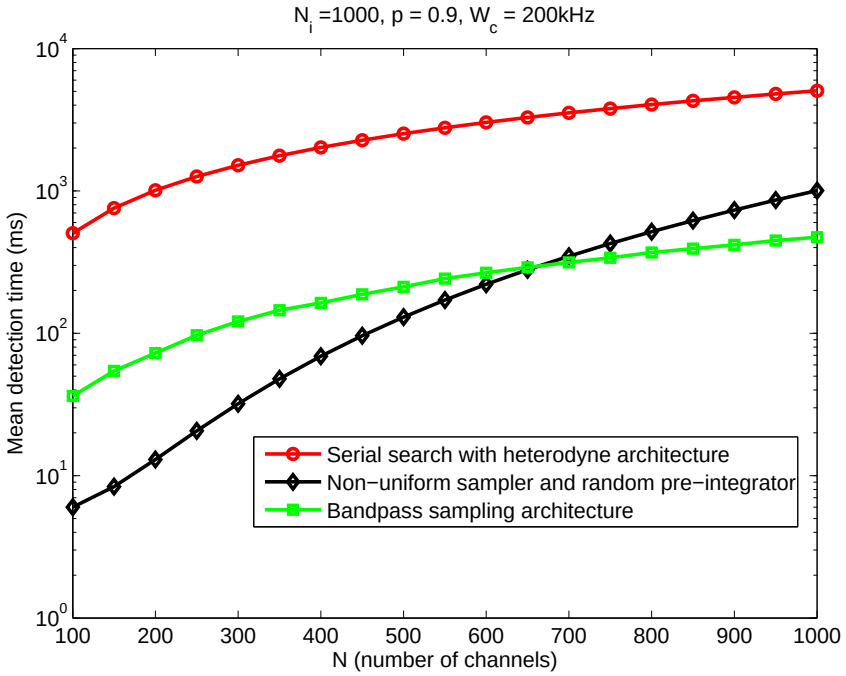
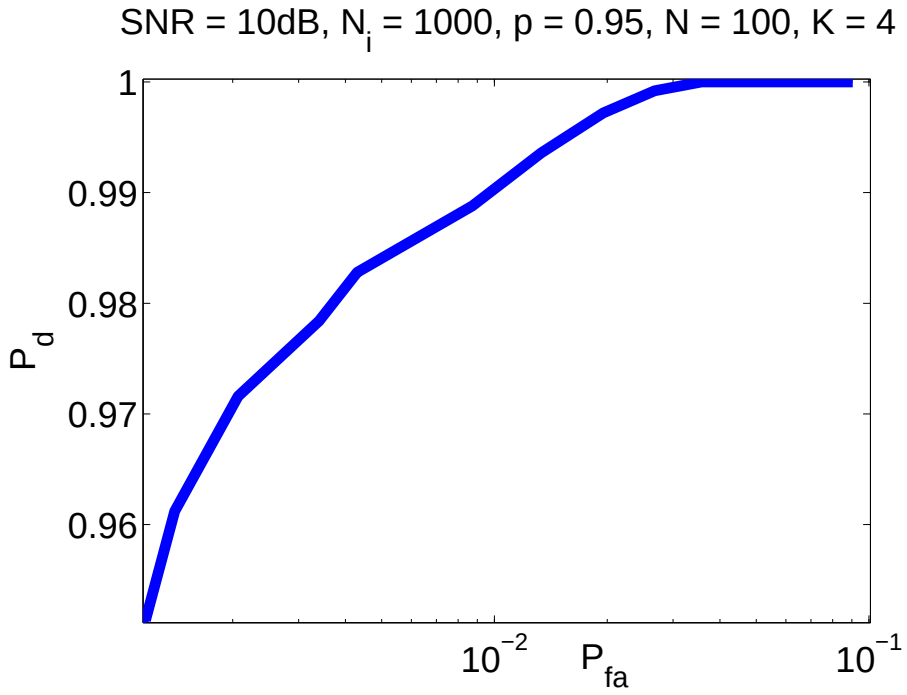
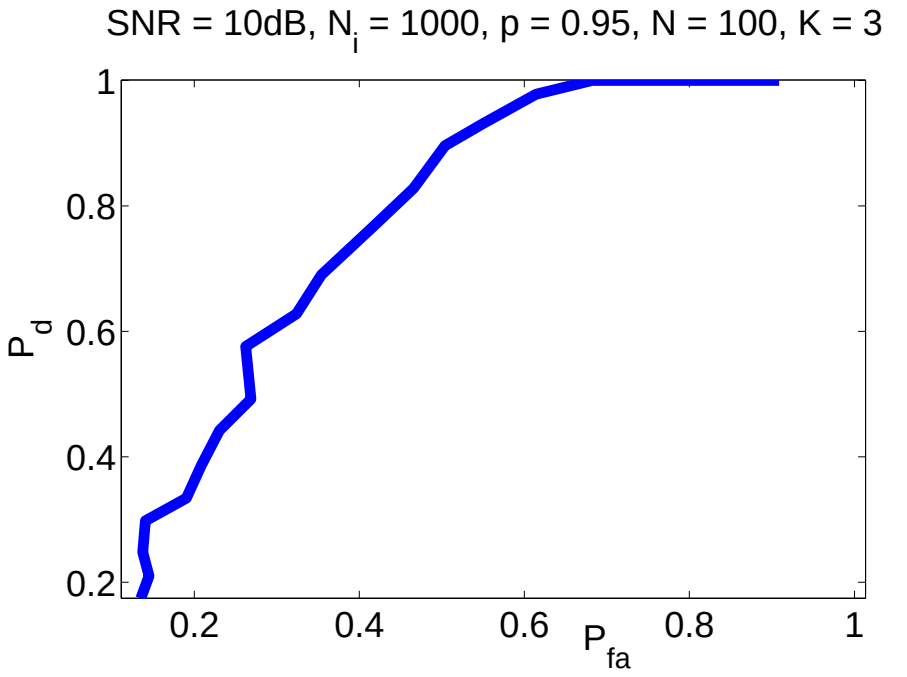


Figure 3.5: Mean detection time of the serial search, non-uniform sampler, random pre-integrator and the proposed bandpass sampling architecture



(a) $K = 4$



(b) $K = 3$

Figure 3.6: Receiver operating characteristics for the proposed bandpass sampling architecture with K sampling rates

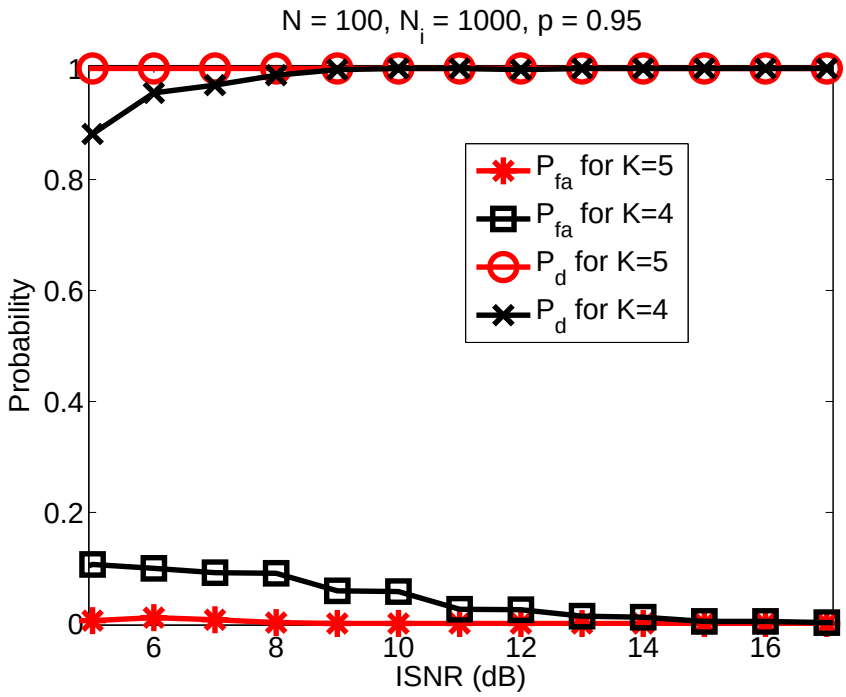


Figure 3.7: Impact of the SNR on the P_d-P_{fa} performance

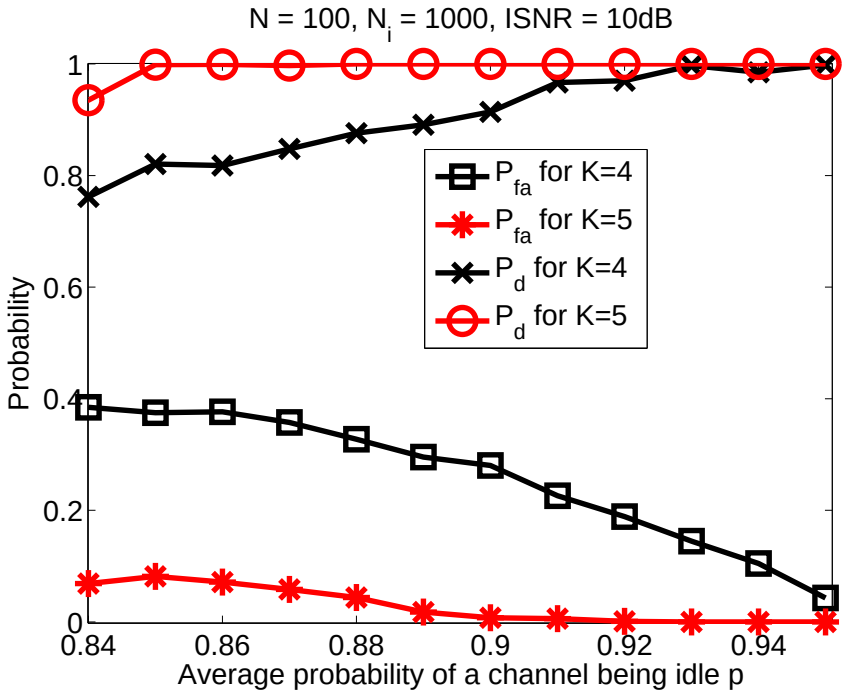
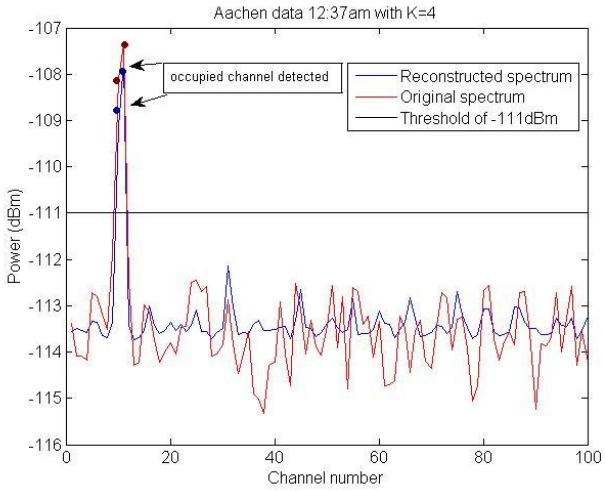
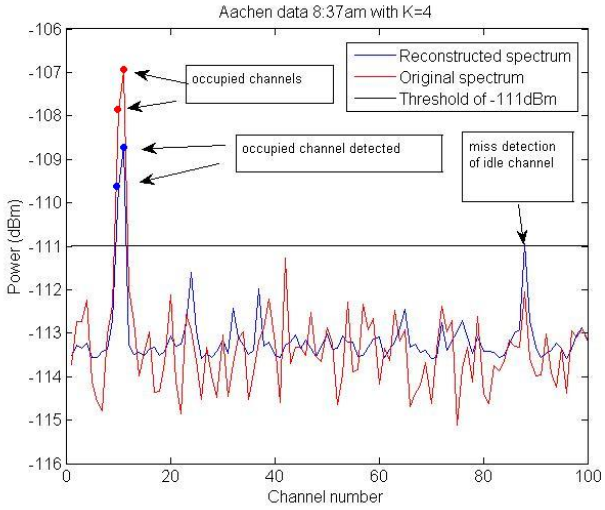


Figure 3.8: Impact of the sparsity level $\bar{p}$ on the P_d-P_{fa} performance

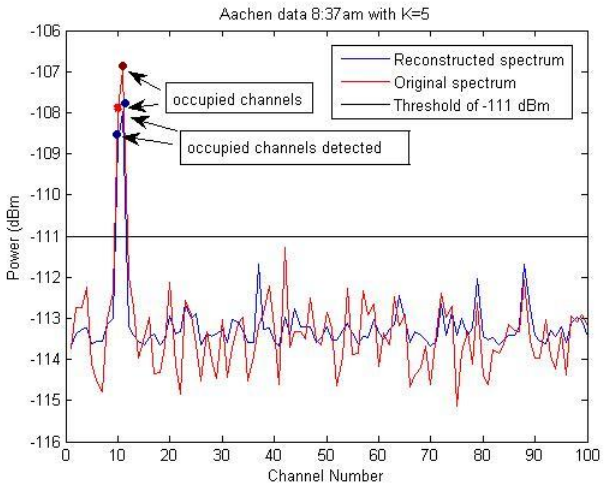


(a) The detection / estimation performance on 12:37am data with K=4



(b) The detection / estimation performance on 8:37am data with K=4

Figure 3.9: Performance of the bandpass sampling architecture on final 100 channels of the Aachen dataset on Jan.27th using K=4



(a) The detection / estimation performance on 8:37am data with K=5

Figure 3.10: Performance of the bandpass sampling architecture on final 100 channels of the Aachen dataset on Jan.27th using K=5

Chapter 4

COMPRESSIVE RADAR CLUTTER SUBSPACE ESTIMATION

4.1 Introduction

In this chapter, we propose the use of dictionary learning algorithm to estimate the clutter subspace, where the angle and frequency axis are discretized *non-uniformly and adaptively* to mitigate the basis mismatch and obtain more accurate estimates of the clutter covariance matrix in STAP. The proposed compressive sensing with dictionary learning (CSDL) algorithm is compared with Multi-stage Wiener filter (MWF) [32], Conjugate Gradient Parametric Adaptive Matched Filter (CGPAMF) [42], Principal Component Inverse (PCI) [83], and Recursive Gram-Schmidt orthonormal basis selection algorithm (RGS). Numerical results based on the Knowledge Aided Sensor Signal Processing and Expert Reasoning (KASSPER) dataset [43] show that CSDL is more accurate than other algorithms with low sample support.

4.2 System Description

4.2.1 System model

Fig. 4.1 shows a typical airborne radar antenna array. Consider a radar consisting of N antenna array elements, each transmitting M coherent pulses at a constant pulse repetition frequency (PRF) in a set range of directions. As given in [85], in a target-free scenario, the received signal is

$$\mathbf{x}_{MN\times 1} = \sum_{i=1}^{N_c} \gamma_i \boldsymbol{\phi}(\theta_i, f_i) + \mathbf{n}_{MN\times 1} = \Phi\boldsymbol{\gamma} + \mathbf{n} \tag{4.1}$$

where

$$\Phi = [\boldsymbol{\phi}(\theta_1, f_1)\ \boldsymbol{\phi}(\theta_2, f_2)\ ...\ \boldsymbol{\phi}(\theta_{N_c}, f_{N_c})] \tag{4.2}$$

and

$$\boldsymbol{\gamma} = [\gamma_1\ \gamma_2\ ...\ \gamma_{N_c}]^T . \tag{4.3}$$

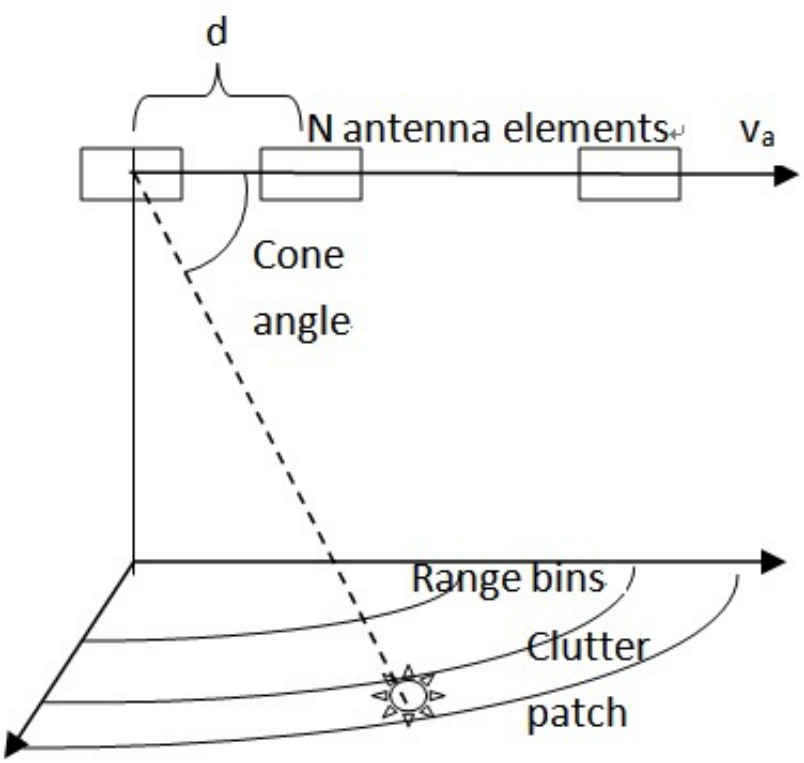


Figure 4.1: A typical STAP scenario

$\mathbf{n} \sim CN(0, \sigma^2\mathbf{I})$ is the additive white Gaussian noise, N_c is the number of clutter patches, θ_i and f_i are the complement of the cone angle and the Doppler frequency of the i^{th} clutter patch, γ_i is the complex scattering intensity (i.e., coefficients with phase) of the i^{th} clutter patch, and

$$\begin{aligned}\boldsymbol{\phi}(\theta_i, f_i) = &\left[1 \; e^{j2\pi\frac{f_i}{PRF}} \; ... \; e^{j2\pi(M-1)\frac{f_i}{PRF}}\right]^T \\ &\otimes \left[1 \; e^{j2\pi\frac{d}{\lambda}sin\theta_i} \; ... \; e^{j2\pi(N-1)\frac{d}{\lambda}sin\theta_i}\right]^T\end{aligned} \tag{4.4}$$

is the steering vector for a given angle and frequency, $\otimes$ is the Kronecker product, d is the distance between two array elements, λ is the radar wavelength.

The optimal STAP matched filter to suppress clutter is then given as in [85]

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{s} \tag{4.5}$$

where $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\}$ is the clutter covariance matrix, and $\mathbf{s}$ is the steering vector for the target. Typically, $\mathbf{R}$ is estimated using the sample matrix inversion (SMI) algorithm with L target-free snapshots, which is given by

$$\mathbf{R}_{SMI} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{x}_i\mathbf{x}_i^H. \tag{4.6}$$

Define the normalized Signal to Interference plus Noise Ratio (SINR) as

$$SINR = \frac{|\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}|^2}{|\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}^H||\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}|} \tag{4.7}$$

where $\hat{\mathbf{R}}$ is the estimated clutter covariance matrix. The normalized SINR shows the degradation of performance due to inaccurate estimate of the clutter covariance matrix compared to knowing the true $\mathbf{R}$. According to the Reed-Mallett-Brennan (RMB) rule [68], a normalized SINR above -3 dB using SMI requires $L \geq 2MN$ snapshots, and such a large number of homogeneous training samples are difficult to obtain in practice. On the other hand, by Brennan's rule [83], $\mathbf{R}$ is a low-rank matrix with rank $r \ll MN$ in most scenarios, leading to reduced-rank processing algorithms, such as PCI, RGS, MWF, CGPAMF, and the proposed CSDL algorithm, which use less snapshots than SMI.

$$\Phi = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_{N_c} \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_{N_c}^{N-1} \\ - - - - - & - - - - - & - - - - - & - - - - - \\ z_1^{\mu} & z_2^{\mu} & \cdots & z_{N_c}^{\mu} \\ z_1^{\mu+1} & z_2^{\mu+1} & \cdots & z_{N_c}^{\mu+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{\mu+N-1} & z_2^{\mu+N-1} & \cdots & z_{N_c}^{\mu+N-1} \\ - - - - - & - - - - - & - - - - - & - - - - - \\ \vdots & \vdots & \vdots & \vdots \\ - - - - - & - - - - - & - - - - - & - - - - - \\ z_1^{(M-1)\mu} & z_2^{(M-1)\mu} & \cdots & z_{N_c}^{(M-1)\mu} \\ z_1^{(M-1)\mu+1} & z_2^{(M-1)\mu+1} & \cdots & z_{N_c}^{(M-1)\mu+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_1^{(M-1)\mu+N-1} & z_2^{(M-1)\mu+N-1} & \cdots & z_{N_c}^{(M-1)\mu+N-1} \end{bmatrix}. \tag{4.17}$$

4.2.2 Sparsity of clutter in frequency-angle domain

As in [83], denote v_a as the speed of the aircraft, then the angle and the Doppler frequency of a static clutter patch satisfy the following relationship, corresponding to a 'clutter ridge'.

$$\frac{f}{PRF} = \mu \frac{d}{\lambda} sin\theta \tag{4.8}$$

where

$$\mu = \frac{2v_a}{d \cdot PRF}. \tag{4.9}$$

The slope of the clutter ridge in frequency-angle domain is thus $\mu \frac{d\ PRF}{\lambda}$.

Denote $z_k = exp(j2\pi \frac{d}{\lambda} sin\theta_k)$. Then for integer μ, Φ in (4.1) can be rewritten as in (4.17). It can be seen from (4.17) that there are only $N + (M-1)\mu$ distinct rows in Φ, while others are repeated rows. For example, if $\mu = 1$, then the row z_1, z_2, ..., z_{N_c} and the row

z_1^μ, z_2^μ, ..., $z_{N_c}^\mu$ are identical, implying that that Φ has a rank of $N + (M-1)\mu$. Therefore, the $N_c \gg N + (M-1)\mu$ clutter patches can be represented in terms of the $N + (M-1)\mu$ basis vectors that span the clutter subspace. Assuming the clutter patches are mutually independent, and independent of additive noise, yields

$$\begin{aligned}
\mathbf{R} &= E\{\mathbf{x}\mathbf{x}^H\} \\
&= E\{(\Phi\boldsymbol{\gamma} + \mathbf{n})(\Phi\boldsymbol{\gamma} + \mathbf{n})^H\} \\
&= \sum_{i=1}^{N_c} E(|\gamma_i|^2)\boldsymbol{\phi}(\theta_i, f_i)\boldsymbol{\phi}^H(\theta_i, f_i) + E\{\mathbf{n}\mathbf{n}^H\} \\
&= \sum_{i=1}^{N_c} E(|\gamma_i|^2)\boldsymbol{\phi}(\theta_i, f_i)\boldsymbol{\phi}^H(\theta_i, f_i) + \sigma^2\mathbf{I} \\
&= \Phi\Gamma\Phi^H + \sigma^2\mathbf{I} \qquad (4.10)
\end{aligned}$$

$$\Gamma = diag(E(|\gamma_1|^2), ..., E(|\gamma_{N_c}|^2)) \qquad (4.11)$$

which means that $\mathbf{R} - \sigma^2\mathbf{I}$ has the same rank as Φ, since $\boldsymbol{\phi}(\theta_i, f_i)$, $i = 1, ..., N_c$ are the columns of Φ. Thus by Brennan's rule, the rank of $\mathbf{R}$ is $r \approx N + (M-1)\mu$ and the measurements $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \sum_{i=1}^{r} \tilde{\gamma}_i \boldsymbol{\phi}(\tilde{\theta}_i, \tilde{f}_i) + \mathbf{n} \qquad (4.12)$$

with the clutter covariance matrix $\mathbf{R}$

$$\mathbf{R} = \sum_{i=1}^{r} E(|\tilde{\gamma}_i|^2)\boldsymbol{\phi}(\tilde{\theta}_i, \tilde{f}_i)\boldsymbol{\phi}^H(\tilde{\theta}_i, \tilde{f}_i) + \sigma^2\mathbf{I} \qquad (4.13)$$

The tilde-sign $\tilde{(\cdot)}$ is used to denote the true basis parameters, where $\boldsymbol{\phi}(\tilde{\theta}_i, \tilde{f}_i)$ and $\tilde{\gamma}_i$ are the basis vectors spanning the clutter subspace and the corresponding complex coefficients, respectively. Equation (4.12) indicates that each measurement can be decomposed into r steering vectors of angle $\tilde{\theta}_i$ and frequency $\tilde{f}_i$, which shows that the clutter is sparse with a rank of $r \ll MN$ in the frequency-angle domain.

4.3 Compressive Sensing With Dictionary Learning

4.3.1 Clutter covariance matrix estimation with compressive sensing

In [72] and [56], CS is applied to estimate the clutter covariance matrix $\mathbf{R}$ by utilizing the sparsity in angle-frequency domain. The angle and frequency axis are discretized into an $N_s \times N_d$ ($N_s = \rho_s N$, $N_d = \rho_d M$) grid, where ρ_s and ρ_d are integers. From (4.1), the received signal is

$$\mathbf{x} = \Phi_{CS}\boldsymbol{\gamma}_{CS} + \mathbf{n} \tag{4.14}$$

where $\boldsymbol{\gamma}_{CS}$ is a sparse vector as discussed in the previous section (since $r \ll N_s N_d$), and Φ_{CS} consists of columns of steering vectors whose angle and frequency corresponds to the intersections in the angle-frequency domain grid. CS theory proves that the sparse vector $\boldsymbol{\gamma}_{CS}$ can be solved from (4.14) by l_1 minimization,

$$\begin{gathered} Minimize\ |\boldsymbol{\gamma}_{CS}|_1 \\ Subject\ to \\ |\mathbf{x} - \Phi_{CS}\boldsymbol{\gamma}_{CS}|_2 < \epsilon \end{gathered} \tag{4.15}$$

where ϵ bounds the resulting noise in the estimation. Equation (4.15) is a convex optimization problem, and can be solved using standard Semi-Definite Programming (SDP) solver in polynomial time. It is proved in [88] that Φ_{CS} in equation (4.14) satisfies RIP. Therefore, given $\boldsymbol{\gamma}_{CS}$ is a sparse vector, it can be uniquely reconstructed. For each non-zero element $\gamma_{i,CS}$ in $\boldsymbol{\gamma}_{CS}$, where i is the index, the i^{th} column in Φ_{CS} is the corresponding estimated basis steering vector $\boldsymbol{\phi}(\theta_{i,CS}, f_{i,CS})$. $\theta_{i,CS}$ and $f_{i,CS}$ can be solved from $\boldsymbol{\phi}(\theta_{i,CS}, f_{i,CS})$, since each steering vector is constructed based on a corresponding grid point $(\theta_{i,CS}, f_{i,CS})$ as in (4.4). Then the clutter covariance matrix $\mathbf{R}$ is calculated from (4.13). With this method, $\mathbf{R}$ can be solved using only $O(r)$ samples as compared to $2MN$ in SMI. However, it requires accurate knowledge of θ_i and f_i to ensure every basis steering vector of the clutter subspace corresponds to an intersection on the grid. Typically, the assumed sparsifying basis is not the same as the true sparsifying basis, resulting in basis mismatch [17], because the representative clutter patches may not located exactly at the grid points. As a result, the

CS reconstruction suffers from an error that grows linearly with the element-wise mismatch between the two basis matrices. Any clutter patch that lies between two grid points leads to a column-wise difference between the true basis matrix and the assumed sparsifying basis matrix, contributing to the reconstruction error.

4.3.2 Dictionary learning

To solve the basis mismatch problem, dictionary learning algorithm is adopted to estimate the difference between the true sparsifying basis matrix and the assumed sparsifying basis matrix in (4.15) by minimize the squared error. Assume for the i^{th} clutter patch, the distance of the true $\tilde{\theta}_i$ and $\tilde{f}_i$ from the nearest grid point is $\Delta\theta_i$ and Δf_i, and the difference between the estimated coefficients via CS and true $\tilde{\boldsymbol{\gamma}} = [\tilde{\gamma}_1 \ ... \ \tilde{\gamma}_{N_s N_d}]$ is $\Delta\boldsymbol{\gamma} = [\Delta\gamma_1 \ ... \ \Delta\gamma_{N_s N_d}]$. If a clutter patch is located at a grid point, the coefficient corresponding to the basis vector in Eq. (4.12) is non-zero; otherwise the entry is zero.

Therefore,

$$\mathbf{x} = \Phi(\theta_{i,CS} + \Delta\theta_i, f_{i,CS} + \Delta f_i)(\boldsymbol{\gamma}_{CS} + \Delta\boldsymbol{\gamma}) + \mathbf{n}. \tag{4.16}$$

$\boldsymbol{\gamma}_{CS} + \Delta\boldsymbol{\gamma}$ is a sparse vector, with only r non-zero elements in it.

Denote $\alpha_i = j2\pi\frac{f_{i,CS}}{PRF}$, $\beta_i = j2\pi\frac{d}{\lambda}sin\theta_{i,CS}$, $\Delta\alpha_i = j2\pi\frac{f_{i,CS}+\Delta f_i}{PRF} - j2\pi\frac{f_{i,CS}}{PRF}$, $\Delta\beta_i = j2\pi\frac{d}{\lambda}sin(\theta_{i_CS} + \Delta\theta_i) - j2\pi\frac{d}{\lambda}sin\theta_{i,CS}$. Then the difference between the columns of the sparsifying matrix used in (4.15) and the true sparsifying matrix is shown in equation (4.18).

Using Taylor expansion $exp(\Delta\beta_i) = 1 + \Delta\beta_i + O(\Delta\beta_i)$, (4.18) can be simplified to Eq. (4.19). Then $\Delta\alpha_i$ and $\Delta\beta_i$ are calculated by minimizing the squared error as in (4.20), where f_g and θ_g are the frequency and angle resolution of the grid used in (4.15), respectively. Equation (4.8) is used as a constraint to guarantee that all the estimated clutter patches are located on the clutter ridge. Since the objective function in (4.20) is a quadratic form of $\Delta\alpha_i$, $\Delta\beta_i$, and $\Delta\gamma_i$, (4.20) can be solved with gradient descent method [41].

The overall CSDL algorithm is summarized in Table 4.1. First, the sparsifying basis matrix and the corresponding complex basis coefficients are estimated with (4.15). Then the dictionary learning algorithm is applied to re-estimate frequency, angle, and basis coefficients

$$\boldsymbol{\phi}(f_{i,CS}+\Delta f_i,\theta_{i,CS}+\Delta\theta_i)-\boldsymbol{\phi}(f_{i,CS},\theta_{i,CS})$$
$$= \begin{bmatrix} 1-1 \\ exp(\beta_i+\Delta\beta_i)-exp(\beta_i) \\ \vdots \\ exp((M-1)(\alpha_i+\Delta\alpha_i)+(N-1)(\beta_i+\Delta\beta_i))-exp((M-1)\alpha_i+(N-1)\beta_i) \end{bmatrix} \quad (4.18)$$

$$\boldsymbol{\phi}(f_{i,CS}+\Delta f_i,\theta_{i,CS}+\Delta\theta_i)-\boldsymbol{\phi}(f_{i,CS},\theta_{i,CS})$$
$$\approx \begin{bmatrix} 0 \\ exp(\beta_i)\Delta\beta_i \\ \vdots \\ exp((M-1)\alpha_i+(N-1)\beta_i)((M-1)\Delta\alpha_i+(N-1)\Delta\beta_i) \end{bmatrix} \quad (4.19)$$

$$\begin{aligned} Minimize\ |\mathbf{x}-\sum_{i=1}^{N_sN_d}(\gamma_{i,CS}+\Delta\gamma_i)\boldsymbol{\phi}(f_{i,CS}+\Delta f_i,\theta_{i,CS}+\Delta\theta_i)|^2 \\ Subject\ to \\ |\Delta f_i|\le f_g \\ |\Delta\theta_i|\le\theta_g \\ \frac{f_{i,CS}+\Delta f_i}{PRF}=\mu\frac{d}{\lambda}sin(\theta_{i,CS}+\Delta\theta_i) \end{aligned} \quad (4.20)$$

---

of the basis clutter patches, and a new sparsifying basis matrix is constructed using (4.17) based on the estimation. After that, (4.15) is executed again. The process continues until the estimated frequency and angle converges. Multiple snapshots are used to estimate the expectation of the power of the clutter coefficients $E(|\tilde{\gamma}_i|^2)$. Then the clutter covariance matrix $\mathbf{R}$ is estimated using Eq. (4.13) and the estimated frequencies, angles, and the expectation of coefficient power.

An example of the objective function for dictionary learning is shown in Fig. 4.2, where

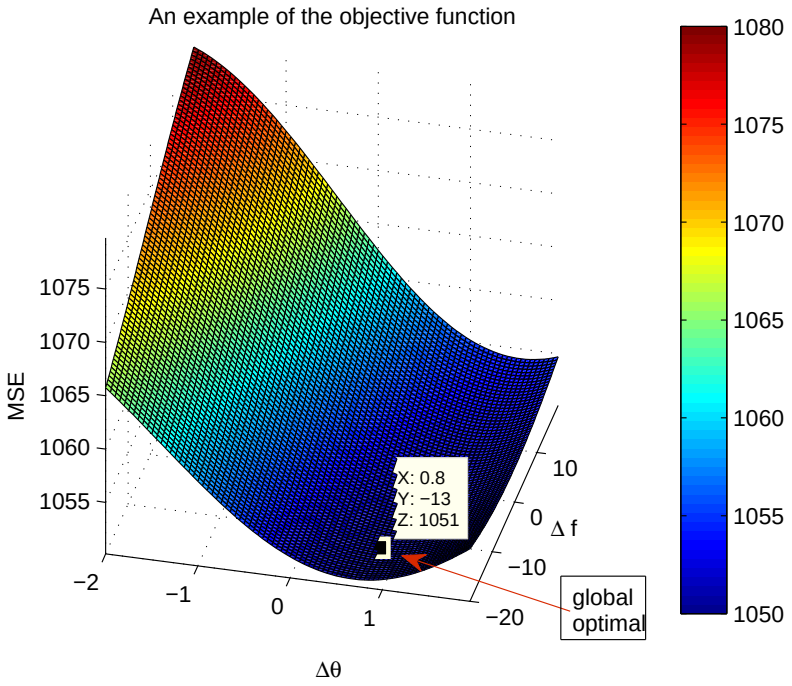


Figure 4.2: Plot of surface of the objective function in Equation (4.20)

the global optimal is marked by an arrow. It can be seen that there are no local optimal that can be reached within the constraints. Therefore, we conclude that the gradient descent method can find the global optimal.

4.4 Numerical Results

The proposed CSDL algorithm is compared with RGS, PCI, MWF, and CGPAMF based on the L-band of the KASSPER dataset [43]. For the KASSPER dataset, the number of antenna elements $N = 11$, and the number of pulses $M = 32$. Therefore, for a normalized $SINR \geq -3$ dB, $L \geq 2MN = 704$ snapshots are needed for SMI method, as discussed in Section 4.2. On the other hand, using the Brennan's rule, $r \approx 42 \ll 704$. This is shown in Fig. 4.3 as a sharp rank cutoff in the eigenspectrum of a clutter covariance matrix in KASSPER dataset. The provided clutter covariance matrix in KASSPER dataset captures the real-world terrain, foliage and urban/manmade structures in a region in CA, USA.

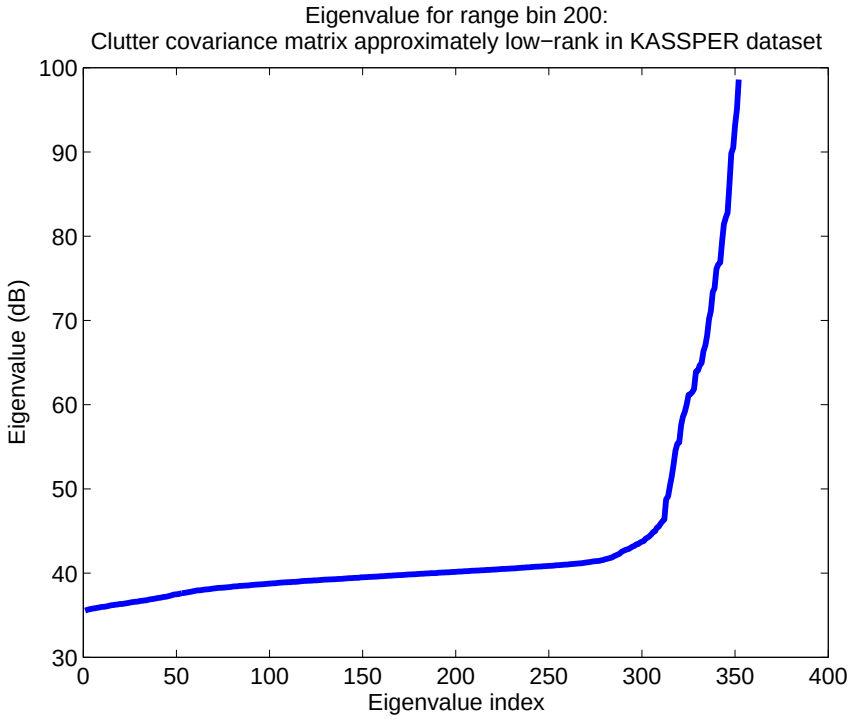


Figure 4.3: Eigenspectrum of KASSPER dataset, range bin 200

Because the true covariance matrices for all the 1000 range bins are available in the dataset, SINR can be computed directly using equation (4.7).

The true covariance matrix for range cell 200 (i.e., the 200^{th} range bin) in KASSPER dateset is used to generate data for multiple snapshots, and additive white noise of clutter-to-noise ratio 40 dB is added to each snapshot. For each snapshot, since the basis coefficients are usually complex Gaussian variables, measurement $\mathbf{x}$ is modeled as colored Gaussian random variables with covariance matrix $\mathbf{R}$ as given in the dataset plus white Gaussian noise. Then the snapshots are processed using the five algorithms. The process is repeated for 10 iterations, where in each of them, new snapshots are generated randomly, and finally the results from the 10 iterations are averaged. This is because in each iteration, both noise and the signal are generated randomly, thus multiple iterations are needed to average out the effects of the randomness, (i.e., to obtain a smooth curve). Comparison between the five algorithms is summarized in Table 4.2.

4.4.1 Computational complexity

The computational complexities, i.e., the operation counts, of the five algorithms are summarized in Table 4.3. It shows that there exists a trade-off between accuracy and computational complexity. CSDL has the highest computational complexity, where $N_d = 352$ and $N_s = 2912$ are used for KASSPER dataset. For the numerical results in this section, $P = 3$ is used for CGPAMF, and $N_{st} = 25$ steering vectors are evaluated.

Using Matlab, the average computation time for processing a snapshot in KASSPER dataset using CSDL is approximately 3 minutes, with an average of 5 iterations for gradient descent algorithm. Note in practice, for processing data on the aircraft, the algorithm is required to converge within 1 ms. Whether the algorithm can be optimized for speed or parallelized is under investigation.

4.4.2 Normalized SINR

The Normalized SINR is calculated as in equation (4.7). It is evaluated as a function of both angle and frequency, since the steering vector $\mathbf{s}$ in (4.7) is a function of angle and frequency. In the figures, SINR is plotted as a function of either angle or frequency by averaging over the other variable. In addition, SINR is plotted in dB, i.e., $SINR_{dB} = 10log_{10}SINR$, and $SINR_{dB} \leq 0$. Note if $\hat{\mathbf{R}} = \mathbf{R}$ for all the frequencies and angles, then SINR is a straight line of $SINR = 0$ dB. As shown in Fig. 4.4 and Fig. 4.5, CSDL has the best performance among all the benchmarks using low sample support (16 samples). The performance of PCI is worse than RGS in this scenario, because r is larger than the number of samples L and PCI uses r basis vectors, leading to inaccurate estimations of $r - L$ basis vectors; whereas RGS uses the basis vectors which capture 99.999% of the covariance matrix energy, thus the number of basis vectors used by RGS is $O(min(L, r))$. The performance for the traditional SMI method is not plotted in the figure, because it is far worse than the five reduced-rank algorithms with low sample support. For $L = 16$, the performance for the traditional SMI method is around -20dB.

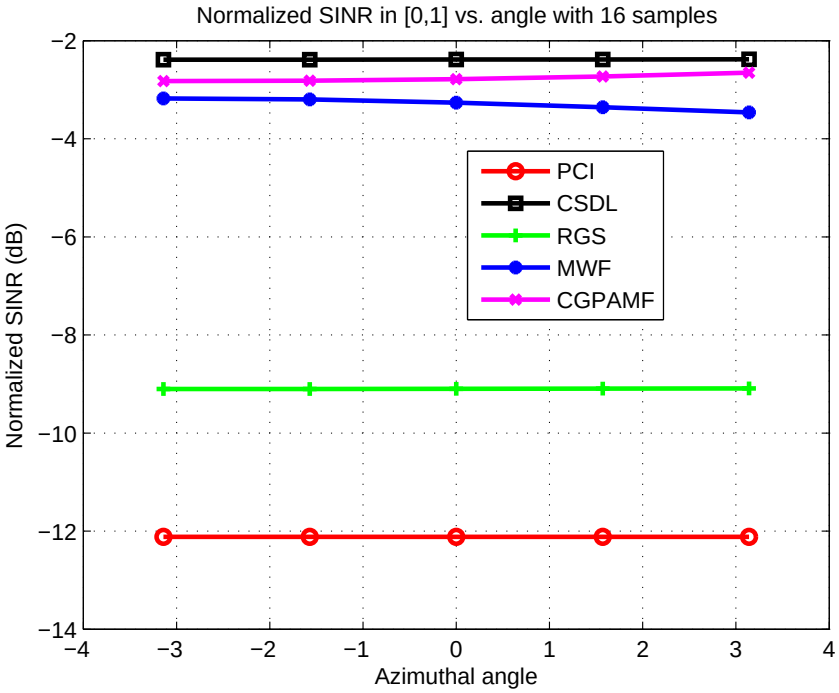


Figure 4.4: Normalized SNR performance versus azimuthal angle

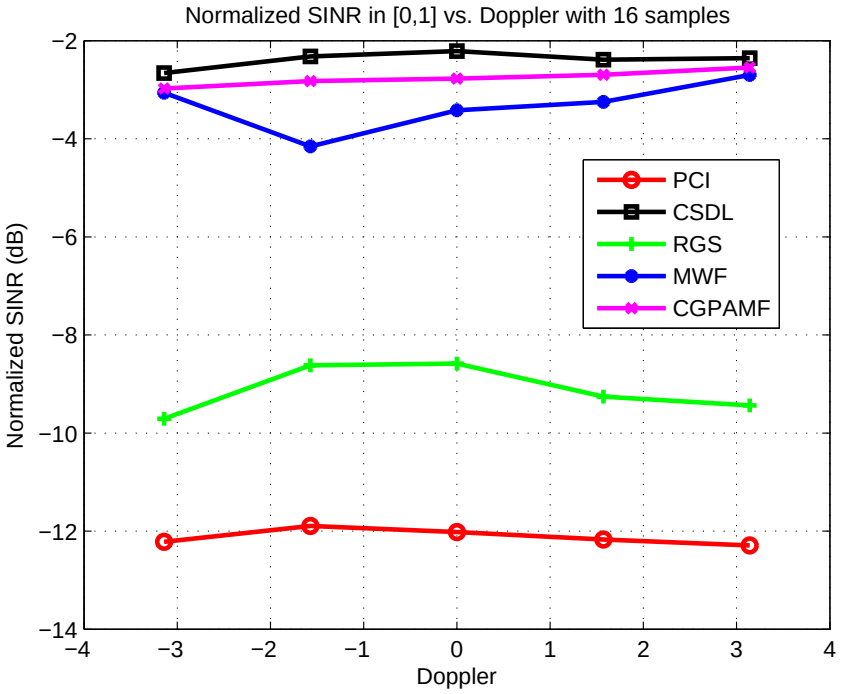


Figure 4.5: Normalized SNR performance versus Doppler frequency

4.5 Conclusion

The CSDL algorithm is proposed to reduce the number of samples needed for clutter subspace estimation in STAP, as well as to mitigate the basis mismatch problem. By exploiting the sparsity of clutter in angle-frequency domain, CSDL performs better than other reduced-rank algorithms, at a cost of high computational complexity.

On the other hand, in practical situations, the clutter structure may not have a sharp rank cutoff, as μ in equation (4.8) may not be an integer. In addition, calibration errors, wind-induced clutter motion and inaccurate measurement of velocity all give rise to the clutter subspace leakage problem. The robustness of the CSDL algorithm against the clutter subspace leakage problem is for future investigation.

Input training data $\mathbf{x}_i$, $i = 1, \ldots, L$
Initialization The sparsifying basis for CS is the grid matrix as in (4.14). $\theta_{i,CS}$, $f_{i,CS}$, and the corresponding $\boldsymbol{\gamma}_{CS}$ are estimated using (4.15).
Stage 1 (a) Re-estimate the angle, frequency, and basis coefficients with (4.20). The sparsifying basis matrix is updated based on the MSE estimation (b) If the angle and frequency from (a) converges, then go to Stage 2. Otherwise, continue to (c) (c) Frequency, angle and basis coefficients are estimated again using (4.15) with the next snapshot. Go back to (a)
Stage 2 Use the estimated sparsifying basis matrix to calculate the basis coefficients $\boldsymbol{\gamma}$ for all the L snapshots, and output the clutter covariance matrix obtained with (4.13)

Table 4.1: CSDL algorithm

Algorithm	RGS	PCI	MWF	CGPAMF	CSDL
Side information	none	Rank r	Rank r, steering vector $\mathbf{s}$	Estimation of the order P	Rank r, clutter ridge slope $\mu\frac{d\ PRF}{\lambda}$
Accuracy	low	low	high	high	high
Computational complexity	low	high	high	high	high

Table 4.2: Comparison between RGS, PCI, MWF, CGPAMF, and CSDL

Algorithm	Complexity
RGS	$O(r^2NM)$
PCI	$O(N^3M^3)$
MWF with N_{st} steering vectors	$O(rN^2M^2N_{st})$
CGPAMF with order P	$O(rN^3Plog_2P)$
CSDL with a grid of $N_d \times N_s$ intersections	$O(rN_d^{1.5}N_s^{1.5}N^2M^2)$

Table 4.3: Computational complexity of the five algorithms

Chapter 5

CONCLUSIONS

In this dissertation, application of compressive sensing to cognitive radio and radar is discussed. We first briefly review the theory of compressive sensing and introduce the background of spectrum sensing in cognitive radio and clutter subspace estimation in cognitive radar.

To detect an idle channel for cognitive radio, a multi-resolution sensing scheme with bandpass sampling is proposed. Bandpass sampling technique enables sub-Nyquist ADC and simplifies the receiver circuits design. To minimize the mean detection time under the constraints of P_d and P_{fa}, the sub-Nyquist sampling factor and the threshold in energy detection are optimized jointly. Numerical results show that the proposed algorithm achieves lower mean detection time compared to one stage algorithms when channel occupancy rate is low.

To detect all the idle channels given the spectrum is sparse, a compressive spectrum sensing architecture with bandpass sampling is proposed. Compared to other compressive sensing architecture, the proposed architecture does not have any Nyquist rate analog circuit components. Numerical results show that the proposed algorithm achieves lower mean detection time compared to the serial search scheme when the channel occupancy rate is low and is a promising alternative when Nyquist rate ADC is not feasible.

Finally, we propose a compressive sensing with dictionary learning algorithm to reduce the number of samples needed for clutter subspace estimation in STAP, as well as to mitigate the basis mismatch problem. By exploiting the sparsity of clutter in angle-frequency domain, CSDL performs better than other reduced-rank algorithms, at a cost of high computational complexity.

BIBLIOGRAPHY

[1] A.A. Abidi. Direct-conversion radio transceivers for digital communications. *Solid-State Circuits, IEEE Journal of*, 30(12):1399 –1410, dec 1995.

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[3] D.M. Akos, M. Stockmaster, J.B.Y. Tsui, and J. Caschera. Direct bandpass sampling of multiple distinct rf signals. *Communications, IEEE Transactions on*, 47(7):983 –988, July 1999.

[4] Ian F. Akyildiz, Won-Yeol Lee, Mehmet C. Vuran, and Shantidev Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks*, 50(13):2127 – 2159, 2006.

[5] Linda Bai and Sumit Roy. A two-stage approach for network monitoring. *Journal of Network and Systems Management*, pages 1–26, 2012.

[6] Y. Bejerano and R. Rastogi. Robust monitoring of link delays and faults in ip networks. *Networking, IEEE/ACM Transactions on*, 14(5):1092 –1103, 2006.

[7] Radu Berinde and Piotr Indyk. Sparse recovery using sparse random matrices. *MIT Technical Report*, Apr 2008.

[8] Richard. P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[9] Tian Bu, Nick Duffield, Francesco Lo Presti, and Don Towsley. Network tomography on general topologies. *SIGMETRICS Perform. Eval. Rev.*, 30(1):21–30, 2002.

[10] M. Ghosh C. Cordeiro, K. Challapali. Cognitive phy and mac layers for dynamic spectrum access and sharing of tv bands. In *Proceedings of IEEE International Workshop on Technology and Policy for Accessing Spectrum*, page 222, August 2006.

[11] Danijela Cabric, Artem Tkachenko, and Robert W. Brodersen. Experimental study of spectrum sensing based on energy detection and network cooperation. In *Proceedings of the first international workshop on Technology and policy for accessing spectrum*, TAPAS '06, New York, NY, USA, 2006. ACM.

[12] Emmanuel J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589 – 592, 2008.

[13] Emmanuel J. Candes and M. B. Wakin. people hearing without listening: an introduction to compressive sampling. *IEEE Signal Processing Magazine*, 2008.

[14] Rui Castro, Mark Coates, Gang Liang, Robert Nowak, and Bin Yu. Network tomography: recent developments. *Statist. Sci.*, 19(3):499–517, 2004.

[15] D. Cavin, Y. Sasson, and A. Schiper. On the accuracy of manet simulators. *Proceedings of the second ACM international workshop on Principles of mobile computing*, pages 38–43, 2002.

[16] Aiyou Chen, Jin Cao, and Tian Bu. Network Tomography: Identifiability and Fourier Domain Estimation. *Proceedings of the IEEE Infocom*, May 2007.

[17] Y. Chi, L.L. Scharf, A. Pezeshki, and A.R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *Signal Processing, IEEE Transactions on*, 59(5):2182 –2195, may 2011.

[18] Wei-Hao Chiu, Yu-Hsiang Huang, and Tsung-Hsien Lin. A dynamic phase error compensation technique for fast-locking phase-locked loops. *Solid-State Circuits, IEEE Journal of*, 45(6):1137 –1149, 2010.

[19] Mark Coates and Robert Nowak. Network loss inference using unicast end-to-end measurement. *Proc. ITC Conf. IP Traffic, Modeling and Management*, pages 28.1–28.9, Sep 2000.

[20] M.A. Davenport, S.R. Schnelle, J.P. Slavinsky, R.G. Baraniuk, M.B. Wakin, and P.T. Boufounos. A wideband compressive radio receiver. In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, 31 2010.

[21] L. Denby, J. M. Landwehr, C. L. Mallows, J. Meloche, J. Tuck, B. Xi, G. Michailidis, and V. N. Nair. Statistical Aspects of the Analysis of Data Networks. *TECHNOMETRICS*, 49(3):318–334, 2007.

[22] Z.H. Derafshi, J. Frounchi, and H. Taghipour. A high speed fpga implementation of a 1024-point complex fft processor. In *Computer and Network Technology (ICCNT), 2010 Second International Conference on*, pages 312 –315, april 2010.

[23] Amogh Dhamdhere, Renata Teixeira, Constantine Dovrolis, and Christophe Diot. Netdiagnoser: troubleshooting network unreachabilities using end-to-end probes and routing data. In *Proceedings of the 2007 ACM CoNEXT conference*, pages 18:1–18:12, New York, NY, USA, 2007. ACM.

[24] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289 –1306, 2006.

[25] Ding-Zhu Du and F. K. Hwang. *COMBINATORIAL GROUP TESTING AND ITS APPLICATIONS*. World Scientific Publishing Company, Singapore, 2000.

[26] Nick Duffield. Network tomography of binary network performance characteristics. *IEEE Trans. Inform. Theory*, 52(12):5373–5388, 2006.

[27] J. P. Elsner, M. Braun, H. Jakel, and F. K. Jondral. Compressed spectrum estimation for cognitive radios. *Proceedings of 19th Virginia Tech Symposium on Wireless Communications*, pages 1–4, June 2009.

[28] H. Fares, M. Ben-romdhane, and C. Rebai. Non uniform sampled signal reconstruction for software defined radio applications. In *Signals, Circuits and Systems, 2008. SCS 2008. 2nd International Conference on*, pages 1 –6, 2008.

[29] Hong Gao, F. K. Hwang, My T. Thai, Weili Wu, and Taieb Znati. Construction of d(h)-disjunct matrix for group testing in hypergraphs. *Journal of Combinatorial Optimization*, 12(3):297–301, 2006.

[30] A. Ghasemi and E.S. Sousa. Collaborative spectrum sensing for opportunistic access in fading environments. In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pages 131 –136, 2005.

[31] Chittabrata Ghosh, Sumit Roy, Marepalli B. Rao, and Dharma P. Agrawal. Spectrum occupancy validation and modeling using real-time measurements. In *Proceedings of the 2010 ACM workshop on Cognitive radio networks*, CoRoNet '10, pages 25–30, New York, NY, USA, 2010. ACM.

[32] J.S. Goldstein, I.S. Reed, and P.A. Zulch. Multistage partially adaptive stap cfar detection algorithm. *Aerospace and Electronic Systems, IEEE Transactions on*, 35(2):645 –661, Apr 1999.

[33] J.R. Guerci and E.J. Baranoski. Knowledge-aided adaptive radar at darpa: an overview. *Signal Processing Magazine, IEEE*, 23(1):41 – 50, Jan. 2006.

[34] Hiroshi Harada and Ramjee Prasad. *Simulation and Software Radio for Mobile Communications*. Artech House, Inc., Norwood, MA, USA, 2002.

[35] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[36] S. Haykin. Cognitive radio: brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2):201 – 220, 2005.

[37] T. Ho, M. Medard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Trans. Information Theory*, 52(10):4413–4430, 2006.

[38] D. Hochbaum. *Approximation Algorithms for NP-hard problems.* PWS Publishing Company, Boston, 1997.

[39] Qian Hui and Yu Lun. Bandpass sampling for multiband signal on exponent demodulation function. In *Mobile Congress (GMC), 2010 Global*, pages 1 –4, 2010.

[40] M.A. Iwen. Simple deterministically constructible rip matrices with sublinear fourier sampling requirements. In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 870 –875, march 2009.

[41] K. Klamroth J. Gorski, F. Pfeuffer. Biconvex sets and optimization with biconvex functions a survey and extensions. *Technical Report*, May 2007.

[42] C. Jiang, H. Li, and M. Rangaswamy. Conjugate gradient parametric detection of multichannel signals. *Aerospace and Electronic Systems, IEEE Transactions on*, 48(2):1521 –1536, Apr. 2012.

[43] J.S.Bergin and P.M.Techau. High-fidelity site-specific radar simulation: Kassper02 workshop datacube. *Information Syst. Laboratories, Inc., Vienna, VA, Tech. Rep. ISL-SCRD-TR-02-105*, 2002.

[44] 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. In *Design, Applications, Integration and Software, 2006 IEEE Dallas/CAS Workshop on*, pages 119 –122, Oct. 2006.

[45] J.N. Laska, S. Kirolos, M.F. Duarte, T.S. Ragheb, R.G. Baraniuk, and Y. Massoud. Theory and implementation of an analog-to-information converter using random demodulation. In *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pages 1959 –1962, may 2007.

[46] I. Loris. On the performance of algorithms for the minimization of l 1 -penalized functionals. *Inverse Problems*, 25(3):035008, 2009.

[47] Ling Luo, N.M. Neihart, S. Roy, and D.J. Allstot. A two-stage sensing technique for dynamic spectrum access. *Wireless Communications, IEEE Transactions on*, 8(6):3028 –3037, 2009.

[48] Ling Luo and Sumit Roy. Analysis of search schemes in cognitive radio. In *Networking Technologies for Software Define Radio Networks, 2007 2nd IEEE Workshop on*, pages 17 –24, 2007.

[49] J. Ma, G.Y. Li, and B. H. Juang. Signal processing in cognitive radio. *Proceedings of the IEEE*, 97(5):805 –823, May 2009.

[50] W.L. Melvin. A stap overview. *Aerospace and Electronic Systems Magazine, IEEE*, 19(1):19 –35, Jan. 2004.

[51] Jia Meng, Wotao Yin, Husheng Li, E. Hossain, and Zhu Han. Collaborative spectrum sensing from sparse observations in cognitive radio networks. *Selected Areas in Communications, IEEE Journal on*, 29(2):327 –337, february 2011.

[52] M. Mishali and Y.C. Eldar. Blind multiband signal reconstruction: Compressed sensing for analog signals. *Signal Processing, IEEE Transactions on*, 57(3):993 –1009, Mar. 2009.

[53] M. Mishali and Y.C. Eldar. Wideband spectrum sensing at sub-nyquist rates [applications corner]. *Signal Processing Magazine, IEEE*, 28(4):102 –135, July 2011.

[54] M. Mishali, Y.C. Eldar, O. Dounaevsky, and E. Shoshan. Xampling: Analog to digital at sub-nyquist rates. *Circuits, Devices Systems, IET*, 5(1):8 –20, 2011.

[55] J. Mitola. The software radio architecture. *Communications Magazine, IEEE*, 33(5):26 –38, may 1995.

[56] H. Morris and M.M. De Pass. Morphological component analysis and stap filters. In *Signals, Systems and Computers, Conference Record of the Forty-First Asilomar Conference on*, pages 2187 –2190, Nov. 2007.

[57] Rajesh Narasimha, Souvik Dihidar, Chuanyi Ji, and Steven W. McLaughlin. Scalable diagnosis in ip networks using path-based measurement and inference: A learning framework. *Journal of Visual Communication and Image Representation*, 21(2):175 – 191, 2010.

[58] H. X. Nguyen and P. Thiran. Active measurement for multiple link failures diagnosis in ip networks. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 185–194, 2004.

[59] H.X. Nguyen and P. Thiran. The boolean solution to the congested ip link location problem: Theory and practice. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 2117 –2125, may 2007.

[60] P.E. Pace, A. Kusmanoff, and G.L. Fudge. Nyquist folding analog-to-information receiver: Autonomous information recovery using quadrature mirror filtering. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pages 1581 –1585, 2009.

[61] V. N. Padmanabhan, L. Qiu, and H. J. Wang. Server-based inference of internet link lossiness. *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03)*, 1:145–155, 2003.

[62] Venkata N. Padmanabhan, Lili Qiu, and Helen J. Wang. Passive network tomography using bayesian inference. *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 93–94, 2002.

[63] S. Pollin, L. Hollevoet, P. Van Wesemael, M. Desmet, A. Bourdoux, E. Lopez, F. Naessens, P. Raghavan, V. Derudder, S. Dupont, and A. Dejonghe. An integrated reconfigurable engine for multi-purpose sensing up to 6 ghz. In *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, pages 656 –657, May 2011.

[64] Yvan Lamelas Polo, Ying Wang, Ashish Pandharipande, and Geert Leus. Compressive wide-band spectrum sensing. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 0:2337–2340, 2009.

[65] M. Rangaswamy. A unified framework for space-time adaptive processing. In *Statistical Signal and Array Processing, 1998. Proceedings., Ninth IEEE SP Workshop on*, pages 360 –363, sep 1998.

[66] J. Ranieri, R. Rovatti, and G. Setti. Compressive sensing of localized signals: Application to analog-to-information conversion. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 30 2010.

[67] B. Razavi. *RF microelectronics*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 1998.

[68] I.S. Reed, J.D. Mallett, and L.E. Brennan. Rapid convergence rate in adaptive arrays. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-10(6):853 –863, nov. 1974.

[69] P. Schniter, L.C. Potter, and J. Ziniel. Fast bayesian matching pursuit. In *Information Theory and Applications Workshop, 2008*, 27 2008.

[70] S. J. Shellhammer. Spectrum sensing in ieee 802.22. *IAPR Wksp. Cognitive Info. Processing*, jun 2008.

[71] The CAIDA Web Site. http://www.caida.org/tools/.

[72] K. Sun, H. Meng, Y. Wang, and X. Wang. Direct data domain stap using sparse representation of clutter spectrum. *Signal Processing*, 91(9):2222 – 2236, Sep. 2011.

[73] Ke Sun, Hao Zhang, Gang Li, Huadong Meng, and Xiqin Wang. A novel stap algorithm using sparse recovery technique. In *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009*, volume 5, pages V–336 –V–339, july 2009.

[74] Frederic Thouin, Mark Coates, and Michael Rabbat. Real-time multi-path tracking of probabilistic available bandwidth. *CoRR*, abs/1010.1524, 2010.

[75] Zhi Tian. Compressed wideband sensing in cooperative cognitive radio networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1 –5, December 2008.

[76] J.A. Tropp and A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655 –4666, 2007.

[77] J.A. Tropp, J.N. Laska, M.F. Duarte, J.K. Romberg, and R.G. Baraniuk. Beyond nyquist: Efficient sampling of sparse bandlimited signals. *Information Theory, IEEE Transactions on*, 56(1):520 –544, 2010.

[78] Yolanda Tsang, Mark Coates, and Robert Nowak. PASSIVE NETWORK TOMOGRAPHY USING EM ALGORITHMS. *IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING*, 6(4031), 2001.

[79] Y. Vardi. Network Tomography: Estimating Source-Destination Traffic Intensities from Link Data. *Journal of the American Statistical Association*, 91(433), 1996.

[80] Vijay V. Vazirani. *Approximation algorithms*. Springer-Verlag, Berlin, 2001.

[81] R.H. Walden. Analog-to-digital converter survey and analysis. *Selected Areas in Communications, IEEE Journal on*, 17(4):539 –550, April 1999.

[82] Yue Wang, Zhi Tian, and Chunyan Feng. A two-step compressed spectrum sensing scheme for wideband cognitive radios. In *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, pages 1 –5, 2010.

[83] J. Ward. Space-time adaptive processing for airborne radar. *Lincoln Laboratory, MIT, Technical Report 10105*, Dec. 1994.

[84] Matthias Wellens and Petri Mhnen. Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model. *Mobile Networks and Applications*, 15:461–474, 2010. 10.1007/s11036-009-0199-9.

[85] M.C. Wicks, M. Rangaswamy, R. Adve, and T.B. Hale. Space-time adaptive processing: a knowledge-based perspective for airborne radar. *Signal Processing Magazine, IEEE*, 23(1):51 – 65, Jan. 2006.

[86] J.J. Wojtiuk. The recovery of carrier envelope information using randomized bandpass sampling. In *Signal Processing and Its Applications, 1999. ISSPA '99. Proceedings of the Fifth International Symposium on*, volume 2, pages 757 –759 vol.2, 1999.

[87] Ye Xia and David Tse. Inference of link delay in communication networks. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 24(12):2235–2248, DEC 2006.

[88] J. Xu and Y. Pi. Compressive sensing in radar high resolution range imaging. *Journal of Computational Information Systems*, 3:778–785, 2011.

[89] A. Y. Yang, A. Ganesh, Z. Zhou, S. Sastry, and Y. Ma. A review of fast l1-minimization algorithms for robust face recognition. *CoRR*, abs/1007.3753, 2010.

[90] Zhuizhuan Yu, S. Hoyos, and B.M. Sadler. Mixed-signal parallel compressed sensing and reception for cognitive radio. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008.

Appendix A

A BRIEF REVIEW OF REDUCED RANK ALGORITHMS FOR CLUTTER SUBSPACE ESTIMATION

A.1 Principal component inverse (PCI)

From the discussions in Chapter 4, the eigenvalue decomposition of $\hat{\mathbf{R}}_{SMI}$ is given by

$$\hat{\mathbf{R}}_{SMI} = \sum_{i=1}^{MN} \lambda_i \mathbf{u}_i \mathbf{u}_i^H \tag{A.1}$$

Denote $\mathbf{u}_i$, $i = 1, ..., r$, as the eigenvectors corresponding to the largest r eigenvalues. As in [83], given a high clutter-to-noise ratio (CNR), the clutter covariance matrix estimation given by PCI is

$$\hat{\mathbf{R}}_{PCI}^{-1} = \mathbf{I} - \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^H \tag{A.2}$$

where $\mathbf{I}$ is the identity matrix.

A.2 Recursive Gram-Schmidt ortho-normal basis selection (Recursive GS)

The recursive GS algorithm is shown in Table A.1. Recursive GS uses Gram-Schmidt process to estimate $\mathbf{u}_i$ in (A.1) instead of eigenvalue decomposition, and then the clutter covariance matrix estimation is obtained by (A.2). Because the ortho-normal basis selected by Recursive GS is within a rotation transformation of the correct subspace, the performance of the recursive GS is worse than PCI with a sufficient sample set, i.e., $L > r$.

A.3 Multi-stage Wiener filter (MWF)

MWF seeks to span the Krylov subspace $\{\mathbf{s}, \mathbf{Rs}, \mathbf{R^2 s}, ..., \mathbf{R^{r-1} s}\}$ instead of eigenvalue decomposition [83]. Since the MWF project the clutter onto the subspace associated with the desired steering vector $\mathbf{s}$, it usually uses less training samples to estimate the clutter subspace. The MWF algorithm is given in Table A.2.

Input: training data $\mathbf{x}_i$, $i = 1, , L$
Recursive Gram-Schmidt process $\mathbf{e}_1 = \mathbf{x}_1$; $\mathbf{e}_1 = \mathbf{e}_1/\lvert\mathbf{e}_1\rvert$ $\mathbf{e}_2 = \mathbf{x}_2 - < \mathbf{e}_1, \mathbf{x}_2 > \mathbf{e}_1$; $\mathbf{e}_2 = \mathbf{e}_2/\lvert\mathbf{e}_2\rvert$ ... $\mathbf{e}_L = \mathbf{x}_L - < \mathbf{e}_1, \mathbf{x}_L > \mathbf{e}_1 - ... - < \mathbf{e}_{L-1}, \mathbf{x}_L > \mathbf{e}_{L-1}$; $\mathbf{e}_L = \mathbf{e}_L/\lvert\mathbf{e}_L\rvert$
Select the basis vectors that corresponds to 99.999% of the matrix energy $\mathbf{u}_1, ..., \mathbf{u}_l$
Calculate and output the estimated clutter covariance matrix $R_{RGS}^{-1} = \mathbf{I} - \sum_{i=1}^{l} \mathbf{u}_i \mathbf{u}_i^H$

Table A.1: Recursive Gram-Schmidt orthonormal basis selection

A.4 Conjugate gradient parametric adaptive matched filter (CG-PAMF)

The CG algorithm is equivalent to the MWF in the matched filter case, since both algorithms seeks to span the same Krylov subspace. For a PAMF, the clutter is assumed to be stationary in time, and the matched filter in Chapter 4 is simplified [42]. The MN-by-1 measurement $\mathbf{x}$ is divided into M equal length vectors $\mathbf{x}(1), ..., \mathbf{x}(M)$, with each N-by-1 vector represents the received signal at the N antenna elements for one of the M pulses. Since the clutter is stationary in time, it can be modeled as a stationary P^{th}-order multichannel AR process

$$\mathbf{x}(n) = \mathbf{B}^H \mathbf{y}_P(n) + \boldsymbol{\epsilon}_P(n) \tag{A.3}$$

where $\mathbf{B}$ is the AR coefficients,

$$\mathbf{y}_P(n) = [\mathbf{x}^T(n-1), \mathbf{x}^T(n-2), ..., \mathbf{x}^T(n-P)]^T \tag{A.4}$$

Equation (A.3) can be break into multiple sub-problems for each of the N antenna elements as

$$\mathbf{x}_j(n) = \mathbf{B}_j^H \mathbf{y}_P(n) + \epsilon_{P,j}(n) \tag{A.5}$$

Then $\mathbf{B}$ is solved from the Wiener-Hopf Equation

$$\mathbf{R}_y \mathbf{B}_j = \mathbf{R}_j \tag{A.6}$$

$\mathbf{R}_y = E[\mathbf{y}_P(n)\mathbf{y}_P^H(n)]$, and $\mathbf{R}_j = E[\mathbf{y}_P(n)\mathbf{x}_j^H(n)]$. Conjugate Gradient algorithm is applied to solve (A.6), then the residue is calculated and the covariance matrix $\mathbf{R}$ is obtained. The CG-PAMF algorithm is summarized in Table A.3.

Input: training data $\mathbf{x}_i$, $i = 1, ..., L$; estimated rank r; steering vector s

Initialization
h(1) = s
X(1) = [x1, x2, x3, , xL]
Bs = 1
 	
Forward recursion
For i = 1:min(r,L)
B(i) = null(h(i)) //null space of h
d(i) = h(i)*a(i) //project onto the target vector
X(i+1) = B(i)*X(i) //project into null space to reduce dimension
rx,d(i) = X(i+1)*d(i) //calculate the next target vector
h(i+1) = rx,d(i)/— rx,d(i)— //normalize the vector
 	
Backward recursion
L(i) = h(i+1)*B(i)*Bs
Ls(i+1) = [Ls(i) ; L(i)]
Bs = B(i)*Bs
End for

Calculate and output the estimated clutter covariance matrix

Table A.2: Multi-stage Wiener filter

Input: training data $\mathbf{x}_i$, $i = 1, ..., L$;

Initialization

$$\mathbf{R}_{xx}^H(m) = \frac{1}{ML}\sum_{k=1}^{L}\sum_{l=1}^{M-m}\mathbf{x}_k(l+m)\mathbf{x}_k^H(l),\ m = -P, ..., P$$

$$\mathbf{R}_y^P = \begin{bmatrix} \mathbf{R}_{xx}(0) & ... & \mathbf{R}_{xx}(P-1) \\ \vdots & ... & \vdots \\ \mathbf{R}_{xx}(1-P) & ... & \mathbf{R}_{xx}(0) \end{bmatrix}$$

$$\mathbf{R}_{yx}^P = \begin{bmatrix} \mathbf{R}_{xx}(-1) \\ \vdots \\ \mathbf{R}_{xx}(-P) \end{bmatrix}$$

Conjugate Gradient algorithm to solve for $\mathbf{B}$

For j = 1 to N

D(1)=-γ(1)=Ryx(j)

B(j)=0

For i = 1 to N(P-1)

$$\alpha(i) = \frac{|\gamma(i)|^2}{D^H(i)R_yD(i)}$$

$$B(j) = B(j) + \alpha(i)D(i)$$

$$\gamma(i+1) = \gamma(i) + \alpha(i)R_yd(i)$$

$$D(i+1) = D(i)\frac{|\gamma(i+1)|^2}{|\gamma(i)|^2} - \gamma(i+1)$$

End for

End for

Calculate the covariance matrix

$$Q = R_{xx}^H(0) - B^HR_{yx}^P$$

$$R_\epsilon = I_{M-P} \otimes Q^{-1}$$

$$C = \begin{bmatrix} B(1) & B(2) & ... & I_N & 0 & ... & 0 \\ 0 & B(1) & B(2) & ... & I_N & ... & 0 \\ & & & \vdots & & & \\ 0 & ... & 0 & B(1) & B(2) & ... & I_N \end{bmatrix}$$

Output $\hat{\mathbf{R}}_{CGPAMF}^{-1} = \mathbf{C}^H\mathbf{R}_\epsilon\mathbf{C}$

Table A.3: Conjugate Gradient Parametric Adaptive Matched Filter algorithm

VITA

Linda Bai received her B.S degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2008, and her M.S. degree in Electrical Engineering from University of Washington, Seattle, WA, USA. She is at present working towards her Ph.D degree in the field of communications and networking. Her current work focuses on applying compressive sensing to network tomography, cognitive radio, cognitive radar, and compressive imaging. She is an IEEE student member. Her Ph.D. thesis is titled Compressive Detection and Estimation with Applications to Cognitive Radio and Radar.

Journal Papers:

Linda Bai and Sumit Roy, "A Two-Stage Approach for Network Monitoring," *Journal of Network and Systems Management*, Vol. 21, Issue 2, pp. 238-263, June 2013.

Linda Bai and Sumit Roy, "Compressive Spectrum Sensing Using a Bandpass Sampling Architecture," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, Vol. 2, Issue 3, pp. 433-442, September 2012.

Conference Papers:

Mohammed H. Firooz, Sumit Roy, Linda Bai, Christopher Lydick, "Link failure monitoring via network coding," *Local Computer Networks (LCN), 2010 IEEE 35th Conference on*, pp. 1068-1075, October 2010.

Linda Bai, Sumit Roy, Muralidhar Rangaswamy, "Compressive Radar Clutter Subspace Estimation Using Dictionary Learning," *IEEE Radar Conference 2013*, Ottawa, Ontario, Canada, May 2013. (Best Student Paper Award Winner)