

Electronic Theses and Dissertations, 2004-2019

2015

Compressive Sensing and Recovery of Structured Sparse Signals

Behzad Shahrabi
University of Central Florida

 Part of the [Electrical and Computer Engineering Commons](#)
Find similar works at: <https://stars.library.ucf.edu/etd>
University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Shahrabi, Behzad, "Compressive Sensing and Recovery of Structured Sparse Signals" (2015). *Electronic Theses and Dissertations, 2004-2019*. 5168.
<https://stars.library.ucf.edu/etd/5168>

COMPRESSIVE SENSING AND RECOVERY OF STRUCTURED SPARSE SIGNALS

by

BEHZAD SHAHRASBI
M.Sc. Oklahoma State University, 2011
B.Sc. Amirkabir University of Technology, 2006

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: Nazanin Rahnavard

© 2015 Behzad Shahrabi

ABSTRACT

In the recent years, numerous disciplines including telecommunications, medical imaging, computational biology, and neuroscience benefited from increasing applications of high dimensional datasets. This calls for efficient ways of data capturing and data processing. Compressive sensing (CS), which is introduced as an efficient sampling (data capturing) method, is addressing this need. It is well-known that the signals, which belong to an ambient high-dimensional space, have much smaller dimensionality in an appropriate domain. CS taps into this principle and dramatically reduces the number of samples that is required to be captured to avoid any distortion in the information content of the data. This reduction in the required number of samples enables many new applications that were previously infeasible using classical sampling techniques.

Most CS-based approaches take advantage of the inherent low-dimensionality in many datasets. They try to determine a sparse representation of the data, in an appropriately chosen basis using only a few significant elements. These approaches make no extra assumptions regarding possible relationships among the significant elements of that basis. In this dissertation, different ways of incorporating the knowledge about such relationships are integrated into the data sampling and the processing schemes.

We first consider the recovery of temporally correlated sparse signals and show that using the time correlation model. The recovery performance can be significantly improved. Next, we modify the sampling process of sparse signals to incorporate the signal structure in a more efficient way. In the image processing application, we show that exploiting the structure information in both signal sampling and signal recovery improves the efficiency of the algorithm. In addition, we show that region-of-interest information can be included in the CS sampling and recovery steps to provide a much better quality for the region-of-interest area compared the rest of the image or video.

In spectrum sensing applications, CS can dramatically improve the sensing efficiency by facilitating the coordination among spectrum sensors. A cluster-based spectrum sensing with coordination among spectrum sensors is proposed for geographically dispersed cognitive radio networks. Further, CS has been exploited in this problem for simultaneous sensing and localization. Having access to this information dramatically facilitates the implementation of advanced communication technologies as required by 5G communication networks.

This dissertation is lovingly dedicated to my mother, Dr. Maryam Khayyat Khameneh, MD. Her support, encouragement, and constant love have sustained me throughout my life.

ACKNOWLEDGMENTS

This dissertation owes its existence to the contributions of many individuals, whom it is my great pleasure to acknowledge. First and foremost, I would like to thank my advisor, Nazanin Rahnavard, for providing an intellectually stimulating and challenging environment in which I conducted this research. Her curiosity, encouragement, engagement, work ethic, persistence, high standards, and patience heavily influenced the way in which I approach research questions, and I am tremendously grateful for the many years of guidance and support she provided me, and the highly accessible manner in which she provided them.

I have been extremely fortunate to be a part of the Communication and Wireless Network Laboratory and enjoy the company of many intelligent, entertaining, and friendly lab-mates, Ali Talari, Minh Tuan Nguyen, Weng Sheng, Ukash Nakarmi, Mohsen Joneidi, and Alireza Zaeemzadeh all provided many hours of insightful, engaging, and oftentimes hilarious conversation and companionship. It was a sincere pleasure to interact with so many diverse people on a daily basis, and to be regularly enlightened, inspired, and gladdened by their presence. Lastly, I thank my family for providing me with the support and encouragement I needed to follow my dreams. Words cannot express my gratitude.

TABLE OF CONTENTS

LIST OF FIGURES	xii
LIST OF TABLES	xix
CHAPTER 1: INTRODUCTION	1
Time-Correlated Compressed Sensing	2
Non-Uniform Compressed Sensing	2
Non-Uniform Sampling and Recovery of Natural Images Using the Hidden Markov Tree Structure of Wavelet Coefficients	3
Coordinated Spectrum Sensing in Cognitive Radio Networks	4
Application of Structured Sparse Compressed Sensing in Cognitive Radio Spectrum Sensing	5
CHAPTER 2: BACKGROUND	6
Compressed Sensing	6
Structured Sparsity	8
Bayesian CS Recovery	9
CHAPTER 3: TC-CSBP: COMPRESSED SENSING FOR TIME-CORRELATED DATA	

	BASED ON BELIEF PROPAGATION	12
Introduction		12
Time-Correlated Data Reconstruction		13
Time-Varying Signal Model		13
CSBP: Compressed Sensing Recovery Using Belief Propagation		18
TC-CSBP: Time-Correlated CS Algorithm Based on Belief Propagation		18
Online Model Parameters Estimation		20
Simulation Results		22
Conclusion		27
CHAPTER 4: NCS: NON-UNIFORM COMPRESSIVE SENSING USING EXPANDER		
	GRAPHS	28
Introduction		28
Contribution of This Chapter		29
Related Work		30
Non-uniform Compressed Sensing (NCS)		32
The Structure of NCS Measurement Matrices		33
The RIP of NCS Measurement Matrices		35

Application of NCS in Burst Image Capture	47
The Structure of the NCS-Based Image Sensor	48
Performance Evaluation	49
Conclusion	60
CHAPTER 5: MODEL-BASED NON-UNIFORM COMPRESSED SAMPLING AND RE- COVERAGE OF NATURAL IMAGES UTILIZING A WAVELET-DOMAIN UNI- VERSAL HIDDEN MARKOV MODEL	61
Introduction	62
Contribution of this chapter	64
The Universal Hidden Markov Tree (uHMT) Model	64
Model-based Non-Uniform Compressed Sampling	70
Non-Uniform Sampling	70
Mutual Coherence of Φ_{uHMT} and Ψ	72
Model-Based Non-Uniform CS Recovery	75
CSBP-uHMT: Integrating the uHMT Model into CSBP	75
AMP-uHMT: Integrating the uHMT into AMP	76
Simulation Results and Discussion	76
Conclusion	84

CHAPTER 6: CLUSTER-CMSS: A CLUSTER-BASED COORIDNATED SPECTRUM

SENSNIG 86

Introduction 86

 Contribution of This Chapter 87

Related Work 88

System Model 90

 The Frame Structure of CRNs 91

 SU’s Belief Vector 92

Cluster-CMSS: Cluster-Based Coordinated Multiband Spectrum Sensing 93

 The Cluster-CMSS Policy With Known PU Dynamic 93

 Learning the PU Dynamic 95

 The Sensing-Based Clustering 99

 Coordinated Spectrum Sensing Within Clusters Using Bipartite Matching 100

Performance Evaluation of Cluster-CMSS 102

 Miss-Detection and False-Alarm probabilities over AWGN channels 102

 Miss-Detection and False-Alarm Probabilities Over Rayleigh and Rician Channels 103

 Energy Cost of Sensing 104

Numerical Results 105

Application of Structured Sparse Compressed Sensing in Cognitive Radio Spectrum	
Sensing	110
Conclusion	111
CHAPTER 7: CONCLUSION	113
Time-Correlated Compressed Sensing	113
Non-Uniform Compressed Sensing	114
Non-Uniform Sampling and Recovery of Natural Images Using the Hidden Markov Tree	
Structure of Wavelet Coefficients	114
Coordinated Spectrum Sensing in Cognitive Radio Networks	115
LIST OF REFERENCES	116

LIST OF FIGURES

Figure 2.1: The non-uniform sparsity of wavelet coefficient of 367 images taken from the 3rd edition of <i>Digital image processing book by Gonzalez and Woods</i> . The wavelet coefficients are divided into 20 bins and the percentage of significant coefficients that lie in each bin are distinguished by the color code.	9
Figure 3.1: Real temperature sensor readings from UC-Berkeley Intel lab. The readings of 54 temperature sensors are considered for 256 consecutive time steps (i.e., every 30 secs). The correlation of the sensor readings in both spatial domain and the time domain can be seen.	15
Figure 3.2: Reconstructed signal at the previous step is fed as <i>a priori</i> knowledge to the decoder.	16
Figure 3.3: Markov model for transition from $X_{i,t-1}$ to $X_{i,t}$	17
Figure 3.4: Block diagram of the learning phase for model parameter estimation.	22
Figure 3.5: Reconstruction LRE versus the number of measurements for TC-CSBP and CSBP algorithms (effect of model mismatch illustration).	24
Figure 3.6: Reconstruction LRE versus the number of measurements for TC-CSBP and CSBP algorithms at $t=10$, for different time correlation parameters.	24

Figure 3.7: Reconstruction LRE in time for the CSBP and Modified-CS algorithms. The signal support only changes at $t = 10$ and it is fixed in all other time slots. The results for two different measurement values (i.e. $m = 75$ and $m = 100$) are given. The simulation parameters that are different than those of Table 3.1 are $n = 250, k = 25$	26
Figure 3.8: Transition probabilities error versus number of times that TC-CSBP algorithm runs ($\xi = 1$).	27
Figure 4.1: The block diagram of the proposed NCS-based image sensor for burst-mode image acquisition.	30
Figure 4.2: Bipartite graph G corresponding to a measurement matrix Φ . Circles and squares represent coefficients \mathbf{x} and measurements \mathbf{y} , respectively, where $\mathbf{y} = \Phi\mathbf{x}$	32
Figure 4.3: The structure of $\Phi_N = [\frac{1}{d_1}\Phi^1 \frac{1}{d_2}\Phi^2 \dots \frac{1}{d_r}\Phi^r]$ in NCS. There are exactly d_l ones randomly placed in each column of Φ^l . Here, $d_1 \geq d_2 \geq \dots \geq d_r$ and a darker color corresponds to a denser matrix.	34
Figure 4.4: The sampling step of NCS. Coefficients in \mathbf{x}_1 and \mathbf{x}_r would have the highest and the lowest presence in the measurements, respectively. The number of variable nodes in section l is $ \mathbf{x}_l = \alpha_l n$ and the degree of a variable node in section l is d_l . The average degree of an encoded node (measurement node) is $L = \frac{n}{m} \sum_{l=1}^r \alpha_l d_l$	35
Figure 4.5: The bipartite graph representation of Φ_N for a toy example with $n = 10, m = 8, r = 2, \alpha_1 = \alpha_2 = 0.5, k_1 = k_2 = 2, d_1 = 4,$ and $d_2 = 2$	39

Figure 4.6: The areas in d_M - d_L plane where (4.16) holds are shown with white color. The design of NCS imposes $d_M \geq d_L$. Therefore the desirable values for d_M and d_L are below the $d_M = d_L$ line. For all d_M and d_L values in those areas, there exists a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph. 47

Figure 4.7: Block diagram representation of the proposed NCS-based image sampling. The rows of non-uniformly sparse Φ_N are used to determine the pixels that their values are multiplexed in analog domain and applied to the input of $\Sigma\Delta$ modulator 49

Figure 4.8: NRE of MICs and LICs versus q_M for *uniformly sparse* signals with $n = 1000$ and $k = 100$. For Φ_N , we set $\alpha = 0.15$, $d = 12$, $m = 300$, and $1 \leq q_M \leq 6$. Note that $q_M = 1$ corresponds to the NRE of uniform CS. 51

Figure 4.9: NRE of MICs and LICs versus q_M , for *non-uniformly sparse* signals with $n = 1000$, $k_1 = 30$, $k_2 = 70$, and $\alpha = 0.15$. For Φ_N we set $\alpha = 0.15$, $d = 12$, $m = 300$, and $1 \leq q_M \leq 6$ 51

Figure 4.10 NRE versus m for $n = 1000$, $\alpha = 0.15$, and $d = 12$, $q_M = 2.5$. The dashed line corresponds to the NRE of uniform CS. 53

Figure 4.11 NRE versus L and $s = \frac{k}{n}$ for MICs and LICs of signals with uniform (solid lines) and non-uniform (dashed lines) sparsities with $n = 1000$, $m = 300$, $\alpha = 0.15$, and $q_M = 2.5$ 54

Figure 4.12A pictorial example for comparing the performances of NCS and uniform CS. Here the image is 128×128 pixels and the number of measurements is $m = 6000$. For NCS, we have $\alpha \simeq 0.1$ and $q_M = 9$. Substantial improvement in the recovery performance of the region of interest (face) is achieved with a slight degradation in the recovery performance of the other parts.	56
Figure 4.13The probability of recovery versus k_1 for our proposed NCS and weighted ℓ_1 [1] for $n = 200$, $m = 100$, $\alpha = 0.5$, and $k_2 = 10$	57
Figure 4.14The average PSNR of 8 consecutive frames that are sampled using proposed NCS. Each frame has 128×128 pixels and is sampled by $m = 6000$ measurements.	58
Figure 4.15Recovered image using NCS and saliency-based CS schemes. The overall PSNR for both schemes is set to be greater than or equal to 25.	59
Figure 5.1: The block diagram of the proposed uHMT-NCS scheme. The vectors \mathbf{x} , \mathbf{y} , and $\hat{\mathbf{x}}$ correspond to the image signal, the CS measurements, and the recovered image, respectively. The uHMT model parameters (Λ) are utilized at both CS Sampling and CS recovery steps.	65
Figure 5.2: The two-dimensional wavelet transform representing an image in terms of approximate coefficients (A), and wavelet coefficients in horizontal (H), vertical (V), and diagonal (D) directions. The wavelet coefficients form quad trees with each parent coefficient having four children in the finer scale. . . .	66
Figure 5.3: The proposed non-uniform measurement matrix for compressive sensing of wavelet coefficients of an image. A darker color corresponds to a denser matrix. 71	

Figure 5.4: Comparison between the recovery performances of our proposed schemes (uHMT-NCS w/ CSBP-uHMT and uHMT-NCS w/ AMP-uHMT) and different CS recovery algorithms (Turbo AMP [2], Model-based CS [3], and TSWCS-MCMC [4]).	77
Figure 5.5: A set of 27 sample images from Microsoft object class recognition database. The numbers in each image are solely for referencing and are not part of the image.	78
Figure 5.6: NRE performance for the sample images in Figure 5.5.	79
Figure 5.7: Comparing the visual performance of different CS schemes at $m = 4000$ measurements.	80
Figure 5.8: Comparing the visual performance of different CS schemes at $m = 6000$ measurements.	81
Figure 5.9: Improvement over the conventional AMP technique by exploiting the uHMT signal model at the sampling step (uHMT-NCS w/ AMP), the recovery step (sparse random sampling w/ AMP-uHMT), and both (uHMT-NCS w/ AMP-uHMT).	82
Figure 5.10 NRE versus L_1 and L_2 . Choosing L_1 and L_2 using (5.9) results in almost 6% performance improvement.	83
Figure 5.11 Recovery performance of uHMT-NCS with CSBP-uHMT recovery versus different depths of the wavelet tree J	84

Figure 6.1: The frame structure of an SU’s operation in a CRN depicting two consecutive time frames. During the sensing time T_S , all SUs cease their transmissions. 92

Figure 6.2: The flowcharts of the proposed Cluster-CMSS policy. The tasks during T_S at the BS and SUs are depicted in right and left boxes, respectively. 94

Figure 6.3: Channel sensing assignment for a network with $N = 6$ (SUs are represented by squares) and $M = 3$ (channels are represented by circles). Subfigure (a) represents the location of SUs. There is a line between two SUs if they are in the transmission range of each other. Subfigure (b) represents the channel sensing assignments and the state of each PU channel on a bipartite graph. In addition the edges with dash-dotted line and solid line represent the failed and the successful sensing attempts, respectively. 96

Figure 6.4: An example of CMSS within a cluster using bipartite matching. 101

Figure 6.5: Average spectrum opportunity discovery ratio versus the PU’s channel utilization λ 106

Figure 6.6: Average spectrum opportunity discovery ratio versus maximum velocity of SU’s movement (v_{max}). 107

Figure 6.7: The characteristic graph (the probability of miss-detection versus the probability of false-alarm) of Cluster-CMSS. Each curve represents different average SNR at the sensing SUs. The performance in the shaded areas are not allowed in the IEEE 802.22 standard. 108

Figure 6.8: The numerically obtained characteristic graph of the proposed Cluster-CMSS policy and the non-cooperative greedy policy in [5] at 15 dB average received SNR. 109

Figure 6.9: The grid of the N_s PU candidate locations. The PUs are present in 2 locations and no other PU is present in the range of those PUs. Therefore the white grid point does not include any PU. Each column represents the activity of PUs in different frequency channel 111

LIST OF TABLES

Table 3.1: Simulation parameters	23
Table 5.1: Properties of different CS sampling and recovery schemes	82
Table 6.1: Average energy cost per successful SU transmission	109

CHAPTER 1: INTRODUCTION

In the year 2013, over 4.4 zettabytes (4.4 trillion GB) of data was created, copied, and transferred [6]. At 40% annual growth rate, this data is estimated to grow to 44 zettabytes by 2020. In the same year, there will be nearly as many digital bits as there are stars in the universe. This volume of data, if stored on 128GB iPad Airs, will require 6.6 stacks of iPads (7.5 mm each) stretching from the Earth to the Moon. This dramatic growth in the size of the data calls for dramatic changes in data acquisition devices and strategies, processing and storage algorithms, and data security.

In many systems such as MRI imaging or hyper spectral imaging, a big portion of the implementation cost is associated with the data acquisition process. For instance in MRI imaging the lengthy process of sample acquisition limits the number of patients that the healthcare provider can serve per day. Likewise the high hardware cost in hyper spectral imagers is a limiting factor. Despite the relatively high cost of sample acquisition in many applications, currently most of the raw data gets compressed almost immediately after acquisition. The more efficient way would be to compress during sampling process. Therefore the redundant information will be avoided at the first place while sampling. In other words, in many real-world problems, we are dealing with high-dimensional data structures such as images (containing many millions of pixels) and there is a high demand to reduce the required measurements for data acquisition.

In what follows, we highlight three research problems that comprise the focus of this dissertation. The first two problems align with the traditional CS focus of sparse signal recovery, while the final problem drifts apart from the domain of typical signal processing problems and applies CS to the localization problem in communication networks. Several common themes can be found across our work on these problems.

Time-Correlated Compressed Sensing

In addition to sparsity, many of the real-world signals have signal components which vary slowly in time. An interesting example is sensor networks in which the signal of interest represents data from temperature sensors that are collected during a time interval T with unit time steps. Such readings have both *spatial correlation* due to closeness of sensors and *time correlation* due to the smooth variations in the temperature. Such a time correlation can further help us to reconstruct the signal at each time step using the estimated signal from the previous time step. In Chapter 3, we propose TC-CSBP, which is a CS recovery algorithm for sparse signals that are also time-correlated. TC-CSBP is based on CS recovery using belief propagation (CSBP) by Baron et al [7]. CSBP serves as the underlying recovery scheme. However, we modify it such that the priori knowledge about the signals coefficients' time correlation is included in the algorithm. Our results show a considerable improvement over conventional CSBP and other related work. Moreover, our results show that TC-CSBP is robust to the error in time-correlation model parameters to a great extent and it can maintain its supremacy in the presence of model mismatch.

Non-Uniform Compressed Sensing

In conventional CS, all signal entries are sampled and recovered *uniformly*. It means that all entries of the signal \mathbf{x} have the same probability of recovery. However, in many real world applications, *non-uniform* acquisition and recovery is desired. For instance, in applications such as medical imaging, computer vision, and geographical information systems (GIS) a particular subset of data may be of more interest relative to the rest of the data. Such subsets are generally known as *region of interest (ROI)*. One might be interested in the recovery of entries associated with ROI with more accuracy. Accordingly, it is very desirable to be able to recover the signal entries non-uniformly.

Such capability is lacking in the conventional CS schemes.

In Chapter 4, we address this problem. We propose NCS for non-uniform sampling of sparse signals. To realize NCS, we employ a *non-uniformly sparse* measurement matrix such that more important coefficients are captured by a relatively larger number of measurements. We analyze the measurement matrix of NCS and show that it satisfies a weaker form of *restricted isometry property (RIP)*. NCS can also be integrated with a recovery algorithm that exploits the non-uniform sparsity of signals [1, 3] for further performance improvement.

Non-Uniform Sampling and Recovery of Natural Images Using the Hidden Markov Tree Structure of Wavelet Coefficients

In Chapter 5, some of the ideas from previous two chapters are combined and are applied to the problem of sampling and recovery of natural images. We modify two Bayesian CS recovery algorithms to incorporate the signal structure (i.e. the Hidden Markov Tree structure of the wavelet coefficients). In addition, based the underlying structure, we propose a measurement matrix that is designed to incorporate the signal structure. Accordingly, we develop a *model-based CS nonuniform sampling and recovery* scheme (uHMT-NCS) for natural images. We show, not only the signal model can be utilized to optimize the initial priors for Bayesian CS recovery algorithms, but it can also be employed in the design of new CS measurement matrices. The results of our numerical experiments suggest a significant performance gain compared to the state-of-the-art model-based CS algorithms.

Coordinated Spectrum Sensing in Cognitive Radio Networks

Cognitive radio (CR) is a promising solution to alleviate today's spectrum deficiency caused by an increased demand for the wireless technologies [8]. The CR paradigm allows a new type of users called unlicensed users or secondary users (SUs) to coexist with the licensed users or primary users (PUs). The SUs are allowed to access the spectrum provided that they do not interfere with the PUs. The under-utilized spectrum bands that can be used by the SUs are called spectrum holes [9].

The availability of spectrum holes varies in both *time* and *space* since the PUs' activity is dispersed in both temporal and spatial domains. An ideal CR is able to efficiently detect and utilize all spectrum holes. Due to the dynamic behavior of PUs, SUs should constantly be aware of the occupancy status of multiple narrow bands or channels of spectrum (a.k.a., wideband spectrum sensing). However, implementing wideband spectrum sensing requires considerable amount of time [10] or complex hardware [11] to obtain a fairly good estimate of the entire spectrum. This lengthy estimation will significantly reduce SUs opportunity to transmit their own data [12].

The problem that we are trying to tackle in Chapter 6 is to develop an spectrum sensing assignment policy that maximizes discovery ratio of spectrum holes while the overhead spectrum sensing time is minimized by sensing only one channel per SU at a time. We propose *cluster-based coordinated multiband spectrum sensing* (Cluster-CMSS). In this problem SUs are mobile and can communicate with a central node or *base station* (BS). This is a very complex problem with numerous challenges. The main challenges are limited ability of SUs in sensing the spectrum, geographically dispersed SU distribution, dynamic PU activity and inaccurate sensing.

A key factor in success of CR networks is environmental awareness. Radio environment map (REM) is a promising tool that provides the environmental awareness for the cognitive radio networks. REM encapsulates multi-domain information from spectrum sensors, geolocation databases, terrain information, and underlying propagation models, and regulatory policies to generate a comprehensive map of the entire spectrum in CR networks [13–15]. REM construction is not an easy task. SUs have to scan a huge swath of spectrum (wideband spectrum sensing) and the PU signal often should be detected in low SNR regime. However to implement the REM, the power spectral density (PSD) information of every point in space at each frequency must be available in real-time.

Building on the spectrum sensing problem that is introduced in this chapter, we will introduce a CS-based spectrum sensing and source localization problem. The localization problem in wireless networks is a well-addressed one [16–18]. However in recent years, with the emergence of cognitive radio networks, has been revisited. Moreover, the application CS-based approaches in PU localization has shown some promises recently [19–21].

CHAPTER 2: BACKGROUND

In this chapter, we briefly present the necessary background on which the dissertation is based. We begin by introducing the compressed sensing. Then, we focus on structured sparsity and the range of problems that show such properties. Finally, we provide background on Bayesian CS recovery.

Compressed Sensing

Emerging *compressed sensing* (a.k.a *compressive sensing*) (CS) techniques [22, 23] provide means to recover a *compressible* signal from its undersampled *random projections* also called *measurements*. Let us define a discrete-time signal of length n as $\mathbf{x} = [x_1, x_2, \dots, x_n]$. The signal \mathbf{x} is said to be k -sparse if it has at most $k \ll n$ non-zero entries in the canonical basis (or in general, k nonzero coefficients in some basis Ψ). The *sparsity rate* of signal \mathbf{x} is defined as $\frac{k}{n}$.

CS paradigm suggests that instead of sampling all the n coefficients of \mathbf{x} , we can recover \mathbf{x} from only $m \ll n$ random measurements [22]. In general, any CS scheme includes the two following key steps:

Signal Sampling: The random projections (measurements) are generated by $\mathbf{y} = \Phi \mathbf{x}$, where $\Phi = [\varphi_{i,j}] \in \mathbb{R}^{m \times n}$ (with $m \ll n$) is a well-chosen random matrix called the *measurement matrix*, and $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector. We can see that the i^{th} measurement is obtained by $y_i = \sum_{j=1}^n \varphi_{i,j} x_j$.

Sparse Signal Recovery: The sparse signal is recovered by obtaining the estimate $\hat{\mathbf{x}}$ from the system of linear equations $\mathbf{y} = \Phi \mathbf{x}$. This is an under-determined system with infinitely many solutions. However, the knowledge of \mathbf{x} being sparse allows us to successfully recover (find a

unique solution) with a high probability from $m = O(k \log n/k)$ measurements by solving the ℓ_1 minimization problem (a.k.a. *basis pursuit* (BP) [24]) given by [22, 24–26]

$$\hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_1, \text{ s.t. } \mathbf{y} = \Phi \mathbf{x}. \quad (2.1)$$

It has been shown that $\hat{\mathbf{x}}$ recovers \mathbf{x} exactly provided that the measurement matrix Φ satisfies a condition known as the *restricted isometry property* (RIP), which is defined as follows,

Definition 1. [23] *A measurement matrix is said to satisfy symmetric form RIP of order S with constant δ_S if δ_S is the smallest number that*

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2, \quad (2.2)$$

holds for every S -sparse \mathbf{x} (i.e. \mathbf{x} contains at most S nonzero entries).

Based on this definition several guarantees are proposed in terms of δ_{2S} , δ_{3S} and δ_{4S} in [27]. In [28] an asymmetric form of Definition 1 is introduced in order to more precisely quantify the RIP.

Definition 2. [28] *For a measurement matrix the asymmetric RIP constants δ_S^L and δ_S^U are defined as,*

$$\begin{aligned} \delta_S^L &= \min_{c>0} (1 - c)\|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2, \\ \delta_S^U &= \min_{c>0} (1 + c)\|\mathbf{x}\|_2^2 \geq \|\Phi \mathbf{x}\|_2^2. \end{aligned} \quad (2.3)$$

Remark 1. [28] *Although both the smallest and largest singular values of $\Phi \Phi^T$ affect the stability of the reconstruction algorithms, the smaller eigenvalue is dominant for compressed sensing in that it allows distinguishing between sparse vectors from their measurements by Φ .*

Let \mathbf{x} be a k -sparse in orthonormal basis $\Psi = [\psi_1, \psi_2, \dots, \psi_n]$ meaning $\boldsymbol{\theta} = \Psi^T \mathbf{x}$ has at most

$k \ll n$ non-zero coefficients. The random measurements are generated by $\mathbf{y} = \mathbf{\Xi}\mathbf{x} = \mathbf{\Phi}\mathbf{\Psi}^T\mathbf{x}$, where $\mathbf{\Phi} = \mathbf{\Xi}\mathbf{\Psi} = [\varphi_{i,j}]$. *Signal recovery* solves $\hat{\mathbf{x}} = \arg \min \|\mathbf{\Psi}^T\mathbf{x}\|_1$ such that $\mathbf{y} = \mathbf{\Xi}\mathbf{x}$.

Structured Sparsity

Conventional CS schemes are oblivious to the distribution of nonzero coefficients in the sparse signal \mathbf{x} . Now the question that may arise is *how we design a new CS scheme with improved performance for the cases in which we have the side information about the signal model*. In this dissertation, we are mostly interested in signals with non-uniform sparsity similar to the non-uniform model considered in [1]. In Chapter 3 the time correlation *structure* between signal coefficient are considered. In Chapters 4 and 5 the *structure* of the wavelet transform coefficients in the images are considered. In Chapter 6 the spectrum sensing problem in cognitive radio networks are being introduced.

While many studies incorporated the signal model or structure at the recovery phase [1, 3, 29–31], only a few papers have looked at the problem from the sampling side [32–35]. It has been shown that structured sparsity (also referred to as *non-uniform sparsity* throughout this dissertation interchangeably) can efficiently be employed in many CS applications such as sensor networks, and image sampling and recovery. As a part of this study in Chapter 4, we show how to incorporate the non-uniform sparsity model into the CS sampling phase.

Non-uniform sparsity can be seen in many practical applications, such as images and videos [36]. Figure 2.1 shows the sparsity pattern of 2D wavelet transform coefficients of a database of 367 images [37] used in the *3rd edition of Digital image processing book by Gonzalez and Woods* [38]. A wide range of images with different sizes are compiled in the book including natural, facial, medical, industrial, and textural images. In Figure 2.1, we have grouped the wavelet coefficients

into 20 bins and have determined the percentage of significant coefficients that lie within each group by a color code. It can clearly be seen from Figure 2.1 that in almost all images the two initial bins contain a great portion of significant coefficients. Therefore the sparsity pattern of almost all images in that database, as it can be seen in Figure 2.1, is non-uniform (i.e. significant coefficients are mostly located at the beginning of the sparse signal and only a few significant coefficients are found at the end of the signal). This non-uniform sparsity can be treated as extra information about the signal.

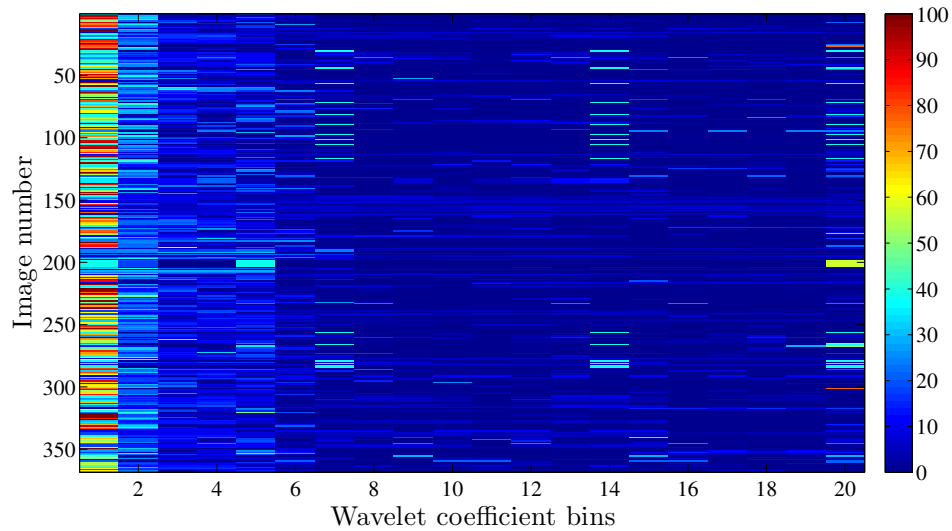


Figure 2.1: The non-uniform sparsity of wavelet coefficient of 367 images taken from the *3rd edition of Digital image processing book by Gonzalez and Woods*. The wavelet coefficients are divided into 20 bins and the percentage of significant coefficients that lie in each bin are distinguished by the color code.

Bayesian CS Recovery

Sparse signal recovery algorithms are divided into three groups. First, those that employ optimization methods such as linear programming [24, 25]. These algorithms usually have a high

computational complexity and might not be practical when n is large. Second, greedy recovery algorithms [39–41], which have lower computational complexities yet require a larger number of measurements to recover the signal compared to the first group. Third, those that are based on Bayesian formulation [7, 42]. Although Bayesian CS recovery is NP complex, it can be efficiently approximated using message passing algorithms [7, 43–46]. Specifically Bayesian compressive sensing via belief propagation (CSBP) [7] and approximate message passing (AMP) [43] are two of the most efficient CS recovery algorithms that employ message passing. CSBP requires a *sparse* measurement matrix to perform efficiently, and AMP can work with the sparse measurement matrices without imposing any penalty on the number of measurements [2]. This is an important property because as we will see in Chapter 5 the measurement matrix itself needs to be *sparse*.

While every method has advantages and disadvantages in different applications, Bayesian approaches represent the most promising potential for the development of structured sparse recovery algorithms. In the literature several publications claimed that Bayesian approaches have a better performance (w.r.t. mean square error (MSE)) compared to non-Bayesian ones [2, 47–49]. The performance of Bayesian approaches greatly depends on how well its signal model is matched to the true signal. Similarly non-Bayesian approaches are sensitive to the statistics of the true signal. In fact, Bayesian methods are more robust to model mismatch compared to non-Bayesian counterparts. Unlike other structured sparsity models [50] where the structure is rigidly enforced and cannot be accommodated all the time, the Bayesian approaches treat the model mismatch as possible but unlikely events. Another advantage of Bayesian approaches is the ability to incorporate graphical models. Graphical model can describe complex signal structures in a transparent way. Graphical models offer ways to capture qualitative description of the signal structure into a rigorous mathematical form. Often Bayesian structured sparsity algorithms rely on a set of parameters. The model parameters can automatically be learned from the data [48].

A major shortcoming of Bayesian approaches is the higher computational cost compared to non-

Bayesian approaches. Typically among greedy algorithms, convex relaxations, and Bayesian approaches, the best performance and slowest running time belong to Bayesian approaches [51]. The high computational cost comes from the inversion of large covariance matrices and/or high number of samples that are needed to approximate the posterior distributions in case of Monte Carlo Markov chain (MCMC) approaches [7]. Our research is motivated by the need for developing accurate and computationally efficient Bayesian algorithms.

CHAPTER 3: TC-CSBP: COMPRESSED SENSING FOR TIME-CORRELATED DATA BASED ON BELIEF PROPAGATION

Existing compressive sensing techniques mostly consider the sparsity of signals in one dimension. However, a very important case that has rarely been studied is when the signal of interest is time varying and signal coefficients have correlation in time. Our proposed algorithm in this chapter is a structure-aware version of the compressive sensing reconstruction via belief propagation proposed by Baron et al. that exploits the time correlation between the signal components and provides the belief propagation algorithm with more accurate initial priors. Numerical simulations show that the belief propagation-based compressive sensing algorithm is able to utilize the side information about signal's time correlation and results in enhanced reconstruction performances.

Introduction

In many signal processing applications, such as image processing and wideband signal processing, an intelligent inspection of signal components reveals that these components are sparse or can be sparsely expressed in a proper basis, e.g. wavelet domain. This observation motivates us to exploit the sparsity of such *compressible signals* to save sampling, communication, processing, and memory resources. These efforts have opened a new area of research known as *compressive (compressed) sensing* (CS) [22, 52, 53]. The underdetermined problem of reconstructing a sparse signal with length n , from its compressed sensed measurement vector with length m , where $m < n$, is possible by using existing CS algorithms [7, 25, 39, 54–57].

In recent years, a considerable amount of research has been conducted to take advantage of the prior knowledge in the reconstruction algorithms [1, 3, 30, 58–64]. The reconstruction algorithms

are mostly the modifications of ℓ_1 reconstruction, except [3, 30, 64] which modify Orthogonal Matching Pursuit [39]. In a recent study [62], CS for the time-correlated signals is considered. However, the proposed scheme requires the collection of all measurements over interval T to reconstruct the coefficients of signals in T . In [1, 58] a weighting strategy is applied to include the prior information into the support of the sparse signal. In other words, the extra information is modulated in terms of different weights for the different parts of the signal support. In [63, 64] the reconstruction of jointly sparse signals in both spatial and temporal domains is considered, and the correlation in spatial domain is employed. In [63], it is assumed that the supports of all correlated signals share an equal common part plus a unique sparse innovation part. In [64], the support is fixed and there is spatial and temporal correlation between signals of the different sources. In this chapter, we assume a different model for the signal than [63, 64].

In [59–61], the authors have developed a novel algorithm that calculates the least-square residuals of the signal supports in two consecutive time steps instead of directly estimating the signal support. In [61] a Kalman filter-based algorithm is introduced to dynamically estimate the residuals.

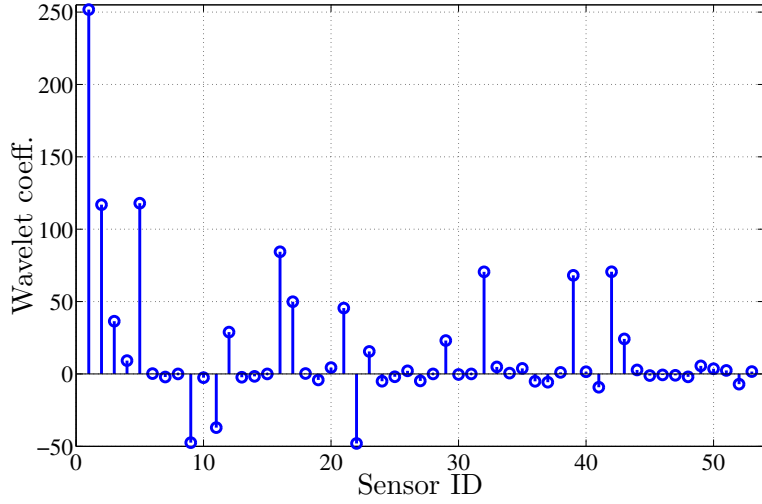
Time-Correlated Data Reconstruction

In this section, we first introduce our time correlation model of signal x_t . Next, we briefly review the CSBP algorithm [7]. Then, we elaborate our approach on incorporating the time correlation in CSBP algorithm.

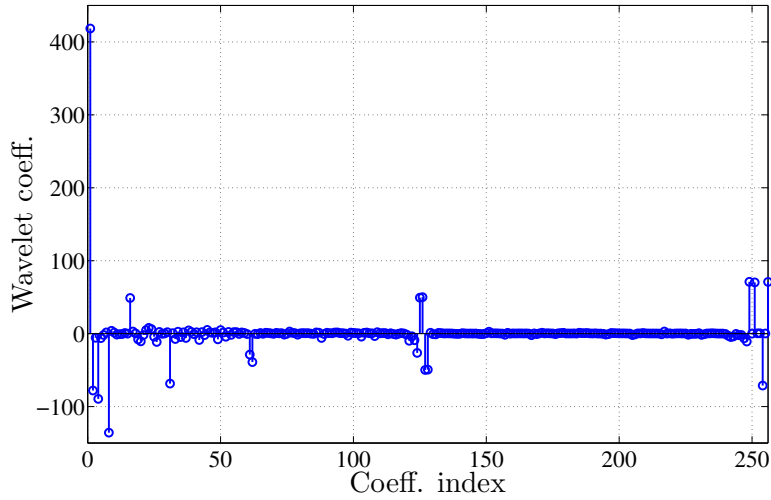
Time-Varying Signal Model

As mentioned, sensor readings in sensor networks represent correlation in both space and time. For example, using the data provided in [65], we depict Fig. 3.1, which shows the correlation of

temperature readings of many thermal sensors in space and time. In Fig. 3.1(a), the readings of 54 sensors at $t = 1$ shows a sparse behavior in wavelet domain (correlation in space). On the other hand, the generated signal from each sensor in time is sparse in wavelet domain (Fig. 3.1(b)) due to time correlation among the sensor readings.



(a) Wavelet coefficients of sensor readings signals ($t = 1$)



(b) Wavelet coefficients of sensor readings in time (Sensor ID= 1)

Figure 3.1: Real temperature sensor readings from UC-Berkeley Intel lab. The readings of 54 temperature sensors are considered for 256 consecutive time steps (i.e., every 30 secs). The correlation of the sensor readings in both spatial domain and the time domain can be seen.

We define $\underline{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{n,t}]$ and $\underline{y}_t = \phi_t \underline{x}_t$ as the signal and CS measurement vectors in time t , respectively. We have $x_{i,t}$ is the value of i^{th} signal coefficient at time t and ϕ_t is the CS

projection matrix at time t . Let $\hat{\underline{x}}_t$ denote the estimate of \underline{x}_t employing a CS recovery scheme. Our goal is to find $\hat{\underline{x}}_t$ exploiting both $\hat{\underline{x}}_{t-1}$ and \underline{y}_t . The information $\hat{\underline{x}}_{t-1}$ is used as *a priori* information in a proper CS decoder to enable the recovery of $\hat{\underline{x}}_t$ with fewer number of measurements as it is shown in Fig 3.2. Only a few of the existing CS algorithms are capable of taking advantage of a priori information about the signal. Among them, we are choosing the state-of-the-art CSBP [7, 57] to implement our proposed ideas.

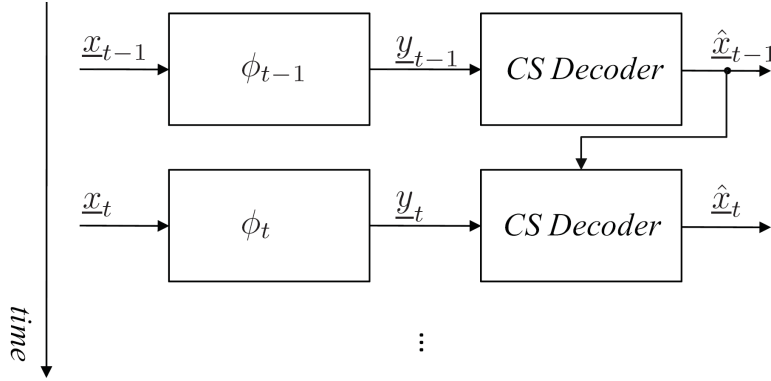


Figure 3.2: Reconstructed signal at the previous step is fed as *a priori* knowledge to the decoder.

Let $\underline{X}_t = [X_{1,t}, X_{2,t}, \dots, X_{n,t}]$ be a random vector and consider \underline{x}_t as an outcome of \underline{X}_t . We know \underline{X}_t and \underline{X}_{t-1} are not independent. Hence, in order to build a mathematical model, we can exploit the correlation between their coefficients (i.e. $X_{i,t-1}$ and $X_{i,t}$ for $i = 1, \dots, n$) in the time domain. We assume \underline{x}_t has only k distinguishable coefficients from the noise level, where $k \ll n$ (k -sparse signal). $\frac{k}{n}$ is defined as the sparsity rate and the value of k can be derived using the history of the signal.

We model variations of each coefficient's value in time by a Markov model (MM) with the state transition matrix $A = \begin{pmatrix} a_{ss} & a_{sl} \\ a_{ls} & a_{ll} \end{pmatrix}$ as shown in Fig. 3.3. According to this model, each element

of vector \underline{X} can be in one of the two states large (\mathcal{L}) and small (\mathcal{S}) that represent whether or not its magnitude is distinguishable from the noise level.

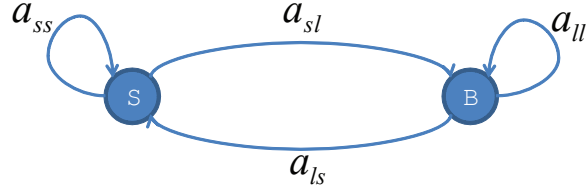


Figure 3.3: Markov model for transition from $X_{i,t-1}$ to $X_{i,t}$.

Initially (at $t = 0$) and similar to [7], the signal coefficients are generated according to the following mixture Gaussian distribution

$$f(X_{i,0}) = \left(\frac{k}{n}\right)\mathcal{N}(0, \sigma_1^2) + \left(1 - \frac{k}{n}\right)\mathcal{N}(0, \sigma_0^2), \quad (3.1)$$

where $\sigma_1 \gg \sigma_0$ and $f(\cdot)$ is the probability density function (pdf). This assumption implies that a fraction $\frac{k}{n}$ of the coefficients of the vector \underline{X}_0 are distributed normally according to $\mathcal{N}(0, \sigma_1^2)$ and the remaining coefficients are distributed according to $\mathcal{N}(0, \sigma_0^2)$. Given (3.1), we define a threshold as $th = 3\sigma_0$ (almost all of the small coefficients lie inside this range), which means that if $\hat{x}_{i,t} > th$ then $\hat{x}_{i,t}$ is in the state \mathcal{L} , otherwise it is in the state \mathcal{S} .

We assume an element in any state at time $t - 1$ is more likely to take its new value in the same state at time step t . This is important knowledge about the variations of the signal in time, which implies that the signal is slowly varying. We model this assumption such that every element in state \mathcal{S} takes a Gaussian distributed value with mean zero and small variance σ_0^2 at each time step, while the coefficients in state \mathcal{L} take a Gaussian distributed value with mean $\hat{x}_{i,t-1}$ and variance σ_0^2 .

In order to maintain a fixed sparsity rate during all of the time steps, $a_{sl} = \frac{k}{n-k}a_{ls}$ should hold. In other words, at each time step on average, the total number of coefficients transiting from \mathcal{S} to \mathcal{L} is equal to the number of coefficients transiting from \mathcal{L} to \mathcal{S} . Otherwise, the signal sparsity changes in time. In the rest of this chapter, we assume that signal sparsity is preserved.

We ignore the effect of noisy measurements on the validity of the model for now. Later, the effect of noisy measurements with standard deviation σ_z is considered. In case of noisy measurements, $\underline{y} = \phi\underline{x} + \underline{\nu}$ where $\underline{\nu} = [\nu_1, \dots, \nu_n]$ is an outcome of random vector \underline{N} and $\nu_i \sim \mathcal{N}(0, \sigma_z^2)$.

CSBP: Compressed Sensing Recovery Using Belief Propagation

The key concept in belief propagation (BP) algorithms is the exchange of beliefs back-and-forth between factor nodes and variable nodes of a factor graph. A factor graph is a bipartite graph in which any vertex from one side of the graph is only connected to the vertices on the other side of the graph [66]. The variable nodes of a factor graph represent the coefficients of \underline{x}_t , and the factor nodes represent randomly generated CS measurements, \underline{y}_t . Moreover, the connecting edges represent which coefficients of the signal vector \underline{x}_t contribute in generating different measurements. The problem is finding the best estimate of each variable node's value using the observations of the factor nodes employing the BP algorithm. CSBP [57] considers the conditional pdf of each element of signal vector \underline{x} as a belief [57], and is especially very interesting since we can employ the prior knowledge of the signal model in terms of a pdf in the reconstruction algorithm.

TC-CSBP: Time-Correlated CS Algorithm Based on Belief Propagation

If we assume the coefficients of vector \underline{x}_t do not show any correlation in time, the only knowledge about the signal is its sparsity in space. Thus, the prior belief about the value of each variable node

at the decoder is in the form of (3.1). Now we consider the case where time-correlation model is added as extra information about the signal to the BP decoding. According to the signal model introduced in system model, we have

$$\begin{aligned} f(X_{i,t}|S(\hat{x}_{i,t-1}) = \mathcal{L}) \\ = a_{ll}\mathcal{N}(\hat{x}_{i,t-1}, \sigma_0^2) + a_{ls}\mathcal{N}(0, \sigma_0^2), \end{aligned} \quad (3.2a)$$

$$\begin{aligned} f(X_{i,t}|S(\hat{x}_{i,t-1}) = \mathcal{S}) \\ = a_{ss}\mathcal{N}(0, \sigma_0^2) + a_{sl}\mathcal{N}(0, \sigma_1^2), \end{aligned} \quad (3.2b)$$

for $i \in \{1, 2, \dots, n\}$ and $S(\cdot)$ represents the state of the signal that can be either large (\mathcal{L}) or small (\mathcal{S}). The proposed TC-CSBP algorithm is different from the original CSBP algorithm [7, 57] in the following key points. First, unlike the conventional CSBP, every variable node receives a *unique a priori belief* according to its previous value ($\hat{x}_{i,t-1}$). Adding this information to the model, the variability of the random vector \underline{X}_t decreases (the first term in RHS of (3.2a) has variance σ_0^2 instead of σ_1^2). Second, in TC-CSBP the time correlation modeling precision is in tradeoff with the number of required measurements. In other words, the more accurate our time correlation model is, the fewer number of measurements is required to achieve a specific reconstruction quality. However, CSBP algorithm's performance only depends on the number of measurements.

Although considering a time correlation model for \underline{x}_t can help the decoder to reconstruct the signal more accurately and with fewer number of measurements, it could also be a source of further errors if the model is not accurate enough. Hence, we need to analyze the robustness of the proposed algorithm to model mismatch and parameter variations. We model these anomalies with random matrix Δ_t that adds up to the state transition probabilities at time step t . Therefore, we face a non-ideal state transition matrix,

$$A_{\Delta} = A + \Delta_t = \begin{pmatrix} a_{ss} + \Delta_s^t & a_{sl} - \Delta_s^t \\ a_{ls} - \Delta_l^t & a_{ll} + \Delta_l^t \end{pmatrix}, \quad (3.3)$$

where $\Delta_t = \begin{pmatrix} +\Delta_s^t & -\Delta_s^t \\ -\Delta_l^t & +\Delta_l^t \end{pmatrix}$ is the anomaly matrix at time t , and Δ_s^t and Δ_l^t are deviations in transition probabilities of states \mathcal{S} and \mathcal{L} , respectively. We consider a time-invariant anomaly matrix (i.e., $\Delta_l^t = \Delta_l$ and $\Delta_s^t = \Delta_s$); however, extension to the time variant case is straightforward. By substituting A with A_{Δ} in (3.2), we obtain

$$\begin{aligned} f(X_{i,t}|S(\hat{x}_{i,t-1}) = \mathcal{L}) &= (a_{ll} + \Delta_l)\mathcal{N}(\hat{x}_{i,t-1}, \sigma_0^2) + (a_{ls} - \Delta_l)\mathcal{N}(0, \sigma_0^2), \\ f(X_{i,t}|S(\hat{x}_{i,t-1}) = \mathcal{S}) &= (a_{ss} + \Delta_s)\mathcal{N}(0, \sigma_0^2) + (a_{sl} - \Delta_s)\mathcal{N}(0, \sigma_1^2). \end{aligned} \quad (3.4)$$

We define a parameter ξ , which we call *model variation to mismatch ratio*, as $\xi_j = \frac{\Delta_j}{\min(a_{js}, a_{jl})}$ for $j \in \{l, s\}$. We assume, this parameter is equal for all rows of a transition matrix (i.e. different states); therefore, expressing the model anomalies using $\xi = \xi_l = \xi_s$.

Online Model Parameters Estimation

The Markov model that is used in previous sections may not be priori known. In this case it is required to employ a learning process along with the signal reconstruction. In such cases, a training data set is required to capture the model. Even after training the algorithm, model parameters mismatch happens in the case of noisy measurements. Measurement noise prevents exact model parameters prediction using the decoded coefficients. Moreover, for non-stationary signals, model parameters gradually change in time. Therefore, in order to avoid anomalies, the model parameters should be periodically estimated using noisy measurements. The offline methods are not of our interest here because they require all decoded signal values in order to predict the model param-

eters. Consequently, we apply an online model parameter estimator such as the one introduced in [67], in our simulations. A *sequential expectation maximization (EM)* algorithm is adapted to estimate the parameters of Markov model (i.e. $\lambda = [A, \sigma_0, \sigma_1]$) sequentially. These sequential algorithms are derived based on maximizing the Kullback-Leibler (KL) information measure, $J(\cdot)$. Given the true model is λ^0 the KL measure between the true model and any model, λ , is defined as $J(\lambda) = E_{\lambda^0} \{\log f(y_t|\lambda)\}$, where $E_{\lambda^0} \{\cdot\}$ is expectation with respect to the true model. The EM algorithm can be summarized as follows,

$$\lambda_t = \arg \max_{\lambda} E_{y_t, \Lambda_t} \{\log f(y_t|\lambda)\}, \quad (3.5)$$

where $\Lambda_t = [\lambda_1, \dots, \lambda_t]$. Using (3.5) the sequential algorithm which is presented in [67, Equations 3.24, 3.30 and 3.33-3.35] can be obtained. This process is fully described in [67].

We update the model parameters only at the end of CSBP iterations at every time step. According to this algorithm, the parameters of Markov model are estimated based on the reconstructed signal coefficient at the previous time steps. Therefore, the algorithm needs to store the previous values of signal coefficients to use them in model parameter estimation. The online parameter estimator architecture can be seen in Fig. 3.4. In Fig. 3.4, the results of signal reconstruction is used in estimating the model parameters A , σ_1 and σ_0 for the next time step.

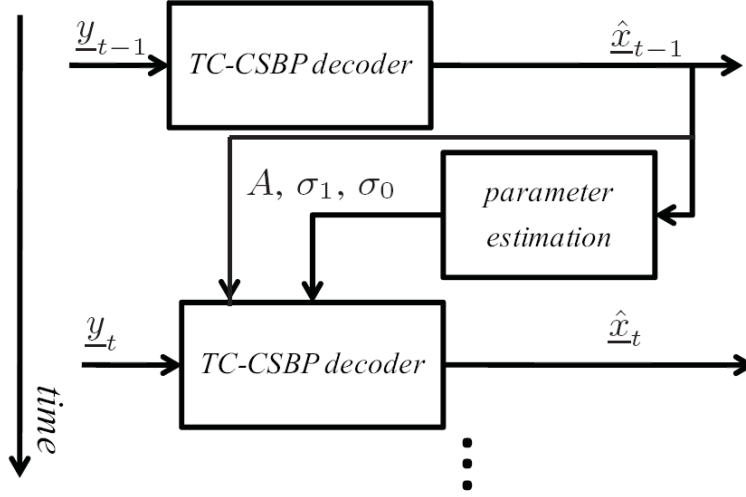


Figure 3.4: Block diagram of the learning phase for model parameter estimation.

Simulation Results

In this section, we evaluate our proposed TC-CSBP method through simulations. The time-correlation model is fed to the proposed TC-CSBP algorithm with and without model mismatch, and then the results are compared with a model-ignorant CSBP algorithm. We also provide the results for comparing the performance of Modified-CS reconstruction algorithm [60] with TC-CSBP algorithm. The numerical simulation parameters are reported in Table 3.1. The simulations of CSBP [68] and modified-CS [60] algorithms are performed using the MATLAB code that the developers provided online. The CSBP algorithm requires a sparse measurement matrix with fixed number of non-zero coefficients at each row and column of ϕ_t . Therefore, we assume a randomly generated measurement matrix with 20 ones at each row. We also assume the model is known at time step $t = 1$ (i.e. the model learning process is finished).

Fig. 6 depicts the ℓ_2 -reconstruction error (LRE), which is defined as $\sqrt{\sum_{i=1}^n (x_{i,t} - \hat{x}_{i,t})^2}$, as a

Table 3.1: Simulation parameters

Parameter name	Value
n : Number of signal samples at each time step	1000
k : Number of coefficients of x_t in state \mathcal{L}	100
σ_1 : Standard deviation of values in state \mathcal{L}	10
σ_0 : Standard deviation of values in state \mathcal{S}	1
σ_z : Standard deviation of observation noise at decoder	1
A : Markov Model State transition matrix	$\begin{pmatrix} 0.989 & 0.011 \\ 0.1 & 0.9 \end{pmatrix}$

function of the number of collected measurements at each time step for CSBP and TC-CSBP with different ξ parameters. It can be seen that TC-CSBP outperforms the conventional CSBP algorithm. For example, for $\xi = 0$ (no model mismatch) and LRE= 30 TC-CSBP requires 400 measurements while CSBP requires about 590 measurements. This means TC-CSBP results in 32% reduction in the number of required measurements. It can also be seen that the proposed algorithm is robust to model mismatch to a good extent. As seen, for $\xi < 0.5$ model parameter variations does not have any destructive effect on the algorithm's performance. On the other hand, it reveals that the proposed algorithm is misled by the model mismatch for relatively large variations in model parameters (i.e., $\xi > 0.5$). Fig. 3.6 illustrates the LRE versus the number of measurement with different signal time correlation parameters. For a highly variable signal, a considerable number of coefficients change their state (i.e., from \mathcal{L} to \mathcal{S} and vice versa) at each time step. As we can see from Fig. 3.6, by increasing the signal variability (i.e. decreasing a_{ll} and a_{ss}), the performance of TC-CSBP algorithm degrades. However, even for a highly variable signal that half of its coefficients at state \mathcal{L} change their state at each time step, TC-CSBP is performs better than conventional CSBP.

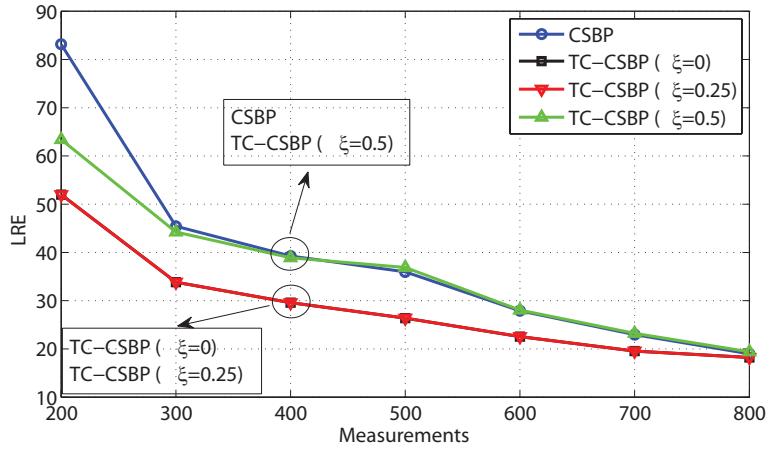


Figure 3.5: Reconstruction LRE versus the number of measurements for TC-CSBP and CSBP algorithms (effect of model mismatch illustration).

In Fig. 6 and Fig. 3.6, the LRE performances of TC-CSBP and CSBP algorithms are measured at the time step $t = 10$.

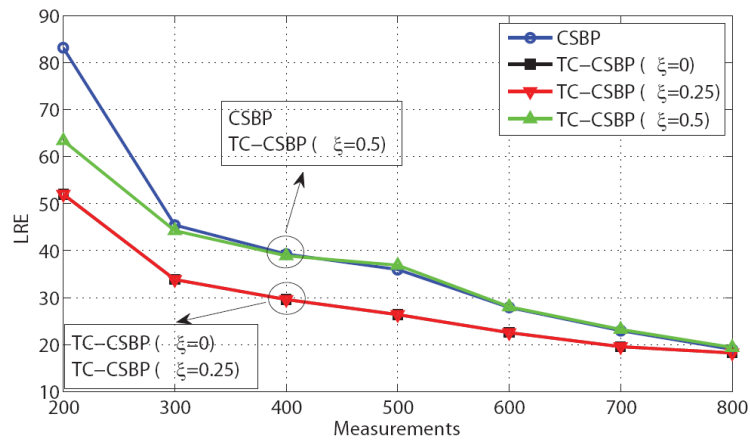


Figure 3.6: Reconstruction LRE versus the number of measurements for TC-CSBP and CSBP algorithms at $t=10$, for different time correlation parameters.

For further comparisons, we have considered a scenario in which the support is only allowed to change at certain time steps. We observe the algorithm's performance over a 20 time step period

and we assume a support change at time step $t = 10$. This change in support consists of adding new coefficients to and removing some existing coefficients from the support, totally 10% of support set changes in $t = 10$. In this setup, the signal coefficients in the state \mathcal{S} are equal to zero and we assume noisy measurements. Using these assumptions, we can compare our algorithm with other studies in the literature [59–61]. Fig. 3.7 shows the reconstruction performances of TC-CSBP and Modified-CS [60] in time with two different values for the number of measurements. At $t = 0$, the reconstruction is performed without using the previously estimated model. The performances of the mentioned algorithms are compared when $m = 75$ and $m = 100$. As we can see, when the number of measurements is small ($m = 75$), TC-CSBP shows a considerable performance improvement, and when an abundance of measurements are available ($m = 100$), Modified-CS (which is based on ℓ_1 -minimization) performs slightly better than TC-CSBP. This is explained by the fact that ℓ_1 reconstruction techniques perform better than CSBP when provided with too many measurements, which is not the case in many CS applications. Modified-CS is the latest version in the series of reconstruction algorithms (e.g. KF-CS [61] and LS-CS [59]) with partially known support.

The robustness of TC-CSBP to error propagation can also be seen from Fig. 3.7. At time step $t = 10$ the support of the signal changes. Consequently, a priori signal values from the previous time step are not valid for the changes in support. As we can see, these types of errors are compensated in a few time steps and do not propagate into the whole algorithm.

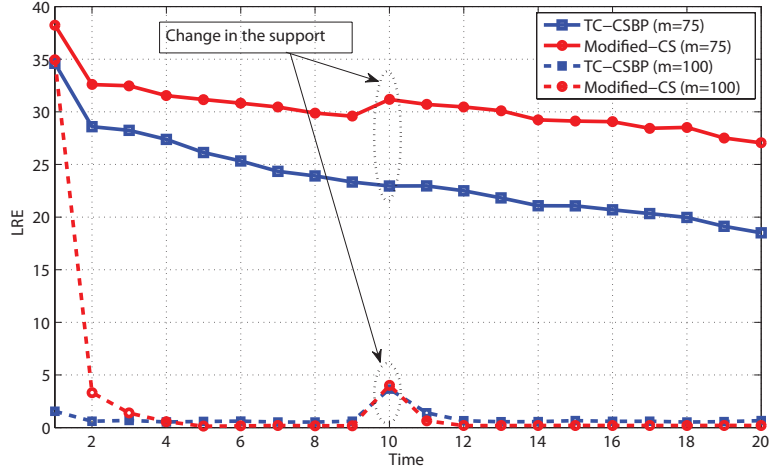


Figure 3.7: Reconstruction LRE in time for the CSBP and Modified-CS algorithms. The signal support only changes at $t = 10$ and it is fixed in all other time slots. The results for two different measurement values (i.e. $m = 75$ and $m = 100$) are given. The simulation parameters that are different than those of Table 3.1 are $n = 250$, $k = 25$.

The learning phase of the online model parameter estimator is shown in Fig. 3.8. We have depicted the absolute error in the estimated transition probabilities versus the number of time steps. It can be inferred from Fig. 3.8 that using the online estimator in 3.5 can compensate the effect of model parameter mismatch after a few time steps.

An interesting trade-off exists between the computational requirements for model parameter estimation and the number of required CS measurements for reaching a given performance. Different designs for this problem can be considered by a system engineer according to the limitations on computation or communication resources. For example, in a multi-hop sensor network, transmitting a symbol over the network is energy-wise more costly than extra computations in one node. On the other hand, implementing a complex decoding algorithm inside a sensor node requires employing a more expensive and more energy consuming CPU.

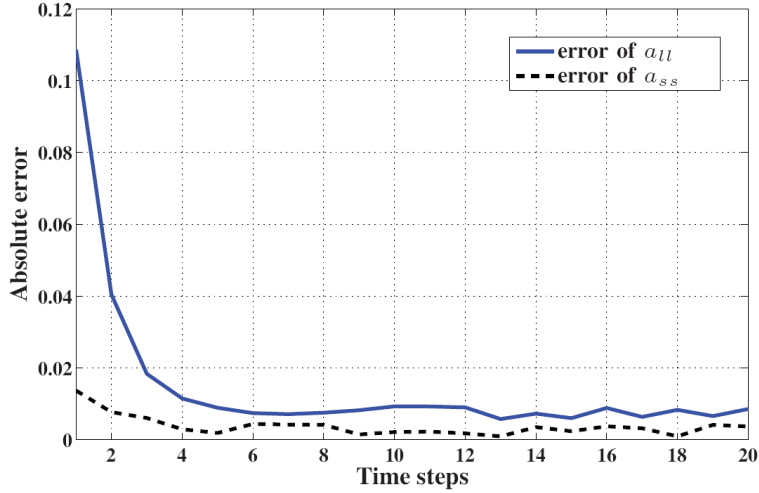


Figure 3.8: Transition probabilities error versus number of times that TC-CSBP algorithm runs ($\xi = 1$).

Conclusion

In this chapter, we have proposed TC-CSBP, which is a compressive sensing reconstruction algorithm for sparse signals that are also time-correlated. TC-CSBP builds upon previous work on compressive sensing via belief propagation (CSBP) by Baron et al. We considered CSBP as our underlying reconstruction scheme due to its flexibility to consider signal model as a priori knowledge. Other CS reconstruction schemes mostly do not have this flexibility. In TC-CSBP, we consider the time-correlation model of signal as a priori knowledge and our results show a considerable improvement over conventional CSBP and other related work. Moreover, our results show that TC-CSBP is robust to the error in time-correlation model parameters to a great extent and it can maintain its supremacy in the presence of model mismatch. Further, an integration of online model estimation into TC-CSBP was studied for more accurate model estimation.

CHAPTER 4: NCS: NON-UNIFORM COMPRESSIVE SENSING USING EXPANDER GRAPHS

In this Chapter, *non-uniform compressive sensing* (NCS) for recovery of sparse signals with either non-uniform sparsity or non-uniform recovery requirement at different parts of the signal. To design NCS, we modify the sampling step of conventional compressive sensing (CS) such that more important coefficients contribute to each CS measurement with a higher probability than less important coefficients. We show that by employing NCS, more important coefficients will be recovered with a lower error rate compared to less important coefficients. To realize NCS, we employ a non-uniformly sparse measurement matrix. We show that the proposed measurement matrix satisfies a weaker version of restricted isometry property (RIP) and we find the sufficient conditions for the existence of such measurement matrices. Next, we perform extensive numerical simulations and show that by correctly setting NCS parameters, the desired non-uniform recovery will be achieved. As an interesting application, we show that NCS can be effectively applied to the burst imaging to provide an enhanced recovery performance for the region of interest (ROI) pixels.

Introduction

Emerging *compressive sensing* (CS) techniques [22, 23] provide means to recover a *compressible* signal from its undersampled *random projections* also called *measurements*. A discrete-time signal $\mathbf{x} = [x_1, x_2, \dots, x_n]$, of length n , is said to be k -sparse if it has at most $k \ll n$ non-zero entries in the canonical basis (or in general, k non-zero coefficients in some basis \mathcal{B}). The *sparsity rate* of signal \mathbf{x} is defined as $\frac{k}{n}$. The CS paradigm suggests that instead of sampling all the n coefficients of \mathbf{x} , we can recover \mathbf{x} from only $m \ll n$ random measurements [22]. The random projections (measurements) are generated by $\mathbf{y} = \Phi \mathbf{x}$, where $\Phi = [\varphi_{i,j}] \in \mathbb{R}^{m \times n}$ (with $m \ll n$) is a well-

chosen random matrix called the *measurement matrix*, and $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector. We can see that the i^{th} measurement is obtained by $y_i = \sum_{j=1}^n \varphi_{i,j} x_j$. The sparse signal is recovered by obtaining the estimate $\hat{\mathbf{x}}$ from the system of linear equations $\mathbf{y} = \Phi \mathbf{x}$. This is an under-determined system with infinitely many solutions. However, the knowledge of \mathbf{x} being sparse allows us to successfully recover (find a unique solution) with a high probability from $m = O(k \log n/k)$ measurements by solving the ℓ_1 minimization problem (a.k.a. *basis pursuit* (BP) [24]) given by [22, 24–26]

$$\hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_1, \text{ s.t. } \mathbf{y} = \Phi \mathbf{x}, \quad (4.1)$$

where $\|\mathbf{x}\|_1 = \sum_i |x_i|$. It has been shown that $\hat{\mathbf{x}}$ recovers \mathbf{x} exactly provided that the measurement matrix Φ satisfies a condition known as the *restricted isometry property* (RIP).

Contribution of This Chapter

An interesting application for our proposed NCS is in the design of CS-based image sensors. Among many proposed techniques in the literature, on-chip realizations with single shot image capture is highly desirable [69]. Employing NCS with these realizations can greatly improve the performance in applications such as high frame rate video and burst image capture. Figure 4.1 shows high-level implementation of NCS in a CS based image sensor. The analog output of the image sensor is applied to a programmable $\Sigma\Delta$ ADC such that it incorporates the non-uniform CS. After this step, the acquired non-uniformly compressed samples are processed to estimate the ROI. The estimated ROI is applied to the $\Sigma\Delta$ ADC to determine the NCS parameters for the acquisition of the next image. This idea is depicted in Figure 4.1 for a burst-mode image capture.

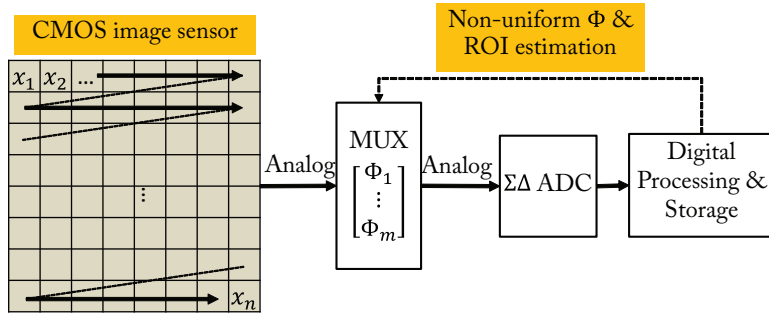


Figure 4.1: The block diagram of the proposed NCS-based image sensor for burst-mode image acquisition.

Related Work

Many existing CS techniques aim at improving the CS recovery performance by exploiting some a priori knowledge about the signal structure. Baraniuk et. al. in [3] proposed to exploit a priori known structure of a signal (for instance the tree structure of signals in wavelet transform) in the recovery step to improve the quality of the recovered signal. He et. al. in [44] proposed to feed the known signal structure to a CS recovery algorithm that is based on Bayesian recovery. Khajehnejad et. al. in [1] proposed weighted ℓ_1 recovery algorithm for non-uniformly sparse signals to improve the recovery performance. The weighted ℓ_1 algorithm penalizes the recovery error of the coefficients in *less* sparse parts of the signal more than the recovery error in *more* sparse parts. Duarte et. al. in [70] designed a *reweighted* basis pursuit recovery algorithm using the structure of the sparse signal in wavelet trees. The aforementioned contributions obtain an improvement by modifying the *recovery step* based on the available *side information* about the signal. In contrast, we modify the *sampling step* by introducing a novel measurement matrix that exploits signal structure.

Our proposed sampling step is based on utilizing sparse measurement matrices. Sparse measurement matrices can be used in lieu of dense matrices to reduce the computational complexity as well as the memory requirements [71]. Gilbert and Indyk in [72] have surveyed different CS recovery algorithms based on the sparse measurement matrices. In [73], the authors have introduced an efficient and low-complexity sparse recovery algorithm for sparse measurement matrices with expander property. In addition, the authors provided a complete comparison of several recovery approaches that use the d -regular expander matrices¹ to create CS measurements.

Other modifications to the structure of the sparse measurement matrices have been proposed in the community. For instance, Kung et. al. in [74] proposed to partition lengthy signals into sections and associated a block of non-zeros in Φ matrix to every section of the signal to reduce the implementation cost. At the recovery step, the similarities among sections were employed to recover the whole signal efficiently. Gan in [32] proposed to have blocks of non-zeros in Φ matrix capturing independent blocks of the signal. However, [32, 74] do not incorporate the knowledge about the *non-uniformity* of the signal sparsity in the sampling step. In CS-based image capture, the application of block sparse measurement matrices is very common [69], mainly because of reduced recovery complexity. In fact a large image is divided into several smaller blocks (i.e., 32×32) and CS is applied to each block individually. In addition, saliency-based compressive sampling scheme in [75] divides the pixels into blocks and assigns different number of measurements to each block based on the number salient pixels in that block. In wireless sensor networks community, a non-uniform data gathering scheme is proposed in [76], in which sensors are selected for data gathering according to a non-uniform Bernoulli model. The measurement matrix in [76] is formed by randomly choosing m out of n rows of the $n \times n$ identity matrix and indeed there is no combination of data in forming CS measurements. In [77], the non-uniform structure of sparse signal is employed to design a block triangular measurement matrix. The structure of measurement matrices in both

¹A d -regular expander matrix is an expander matrix with exactly d non-zeros per column.

[76, 77] is completely different than what we propose here. NCS is mainly inspired by a previous work by Rahnavard et al. on *unequal error protection* rateless codes [78].

Non-uniform Compressed Sensing (NCS)

Let \mathbf{x} be a *non-uniformly* sparse signal of length n , partitioned into r uniformly-sparse sections $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ (i.e., $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r]^T$) with sizes $\alpha_1 n, \alpha_2 n, \dots, \alpha_r n$, respectively, with $\sum_{l=1}^r \alpha_l = 1$, and *decreasing* sparsity rates (*decreasing* importance levels). The number of non-zero entries in \mathbf{x}_l is at most k_l , for $l = 1, 2, \dots, r$. We refer to such \mathbf{x} as an $(\boldsymbol{\alpha}, \boldsymbol{\kappa})$ -sparse signal, where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_r]$ and $\boldsymbol{\kappa} = [k_1, \dots, k_r]$.

Similar to [7], we may view the coefficients \mathbf{x} and the CS measurements $\mathbf{y} = \Phi \mathbf{x}$ as vertices of a bipartite graph G , where all signal coefficients are *variable* nodes and the measurements are *encoded* nodes. The coefficient x_j is connected to the measurement y_i with an edge of weight $\varphi_{i,j}$ (see Figure 4.2).

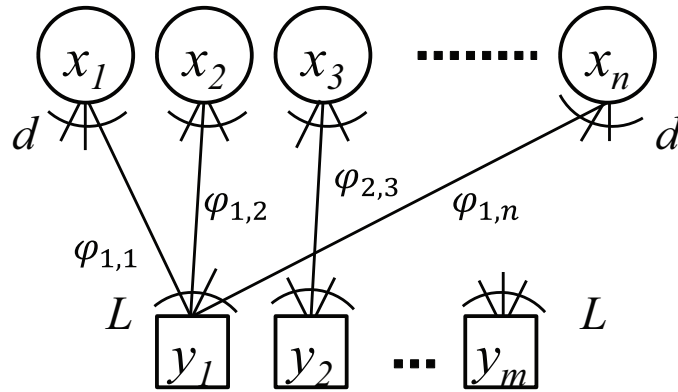


Figure 4.2: Bipartite graph G corresponding to a measurement matrix Φ . Circles and squares represent coefficients \mathbf{x} and measurements \mathbf{y} , respectively, where $\mathbf{y} = \Phi \mathbf{x}$.

Conventionally, Φ matrices are chosen to be dense with iid Gaussian or Bernoulli entries. However, employing dense measurement matrices requires high computational complexity and a large memory. Instead, in applications with restrictions on these resources, it is desirable to use sparse measurement matrices [71–73]. In all these contributions Φ provides *uniform* sampling and *uniform* recovery for all coefficients. Figure 4.2 shows G when Φ is a *uniformly* sparse matrix, where all coefficients have degrees (the number of edges connected to a node) equal to d and all measurements would have degree L , where $dn = Lm$.

On the contrary, we design a *non-uniformly* sparse measurement matrix at the sampling step, and we propose to incorporate more important signal coefficients in a larger number of measurements, or equivalently sample them with a higher frequency. This leads to a non-uniform distribution of edges over variable nodes in G .

The Structure of NCS Measurement Matrices

Our proposed measurement matrix $\Phi_N = [\frac{1}{d_1}\Phi^1 | \frac{1}{d_2}\Phi^2 | \dots | \frac{1}{d_r}\Phi^r]$ is composed of r submatrices Φ^l of sizes $m \times \alpha_l n$ that are scaled by $\frac{1}{d_l}$, for $l = 1, 2, \dots, r$ so that every column of Φ_N adds up to 1. Each column of the submatrix Φ^l , which corresponds to a signal coefficient in \mathbf{x}_l , is designed to have exactly d_l ones at random positions. Therefore, using Φ_N , a coefficient in \mathbf{x}_l only contributes to d_l measurements. A higher value of d_l maps to a higher density of ones in Φ^l . We desire more important coefficients to contribute to a larger number of measurements; therefore, we set $d_1 \geq d_2 \geq \dots \geq d_r$. Figure 4.3 shows the structure of Φ_N , which enables the non-uniform sampling. By setting the appropriate values for d_l 's, we can adjust the frequency that measurements capture the coefficients from \mathbf{x}_l 's.

$$\Phi_N = \left[\begin{array}{c|c|c} \frac{1}{d_1} \Phi^1 & \frac{1}{d_2} \Phi^2 & \dots & \frac{1}{d_r} \Phi^r \\ \hline \alpha_1 n & \alpha_2 n & & \alpha_r n \end{array} \right]_{m \times n}$$

Figure 4.3: The structure of $\Phi_N = [\frac{1}{d_1} \Phi^1 | \frac{1}{d_2} \Phi^2 | \dots | \frac{1}{d_r} \Phi^r]$ in NCS. There are exactly d_l ones randomly placed in each column of Φ^l . Here, $d_1 \geq d_2 \geq \dots \geq d_r$ and a darker color corresponds to a denser matrix.

An ensemble of NCS is specified by parameters $n, m, \alpha = \{\alpha_1, \dots, \alpha_r\}$, and $\mathbf{d} = \{d_1, \dots, d_r\}$. Figure 4.4 depicts the bipartite graph representation of the proposed Φ_N .

It is straightforward to show that the number of non-zero entries in a row of Φ_N (the degree of a measurement) is a random variable with mean $L = \frac{n}{m} \sum_{l=1}^r \alpha_l d_l$.

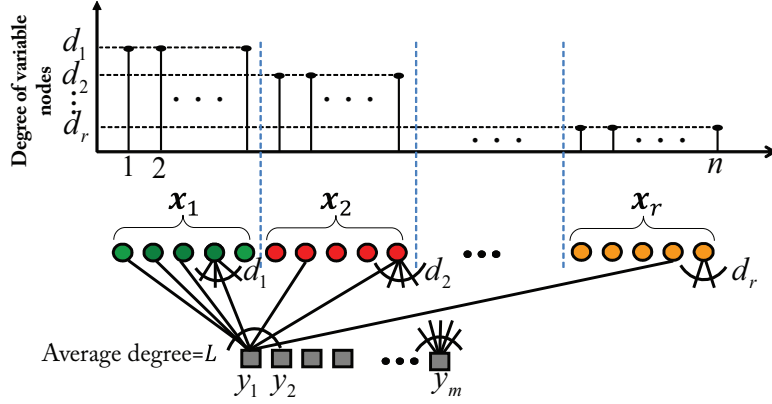


Figure 4.4: The sampling step of NCS. Coefficients in \mathbf{x}_1 and \mathbf{x}_r would have the highest and the lowest presence in the measurements, respectively. The number of variable nodes in section l is $|\mathbf{x}_l| = \alpha_l n$ and the degree of a variable node in section l is d_l . The average degree of an encoded node (measurement node) is $L = \frac{n}{m} \sum_{l=1}^r \alpha_l d_l$.

The RIP of NCS Measurement Matrices

The question that arises here is whether employing Φ_N satisfies CS requirements similar to a dense Φ . Originally, it was proven that ℓ_1 -minimization-based approaches find the unique solution to the under-determined CS problem, if the measurement matrix satisfies the RIP(2) (see Definition 3) condition [23]. However, it was later shown that the weaker RIP(p) condition, with $p \geq 1$, yields to an accurate solution as well [71, 72]. Sparse Φ matrices do not satisfy the RIP(2) condition, unless their number of rows is large [72]. However, it has been shown that if a sparse Φ is the adjacency matrix of a (k, d, ϵ) -regular expander graph (see Definition 4), it satisfies RIP(1) condition [72]. Hence, the sparse measurement matrix Φ can be used to efficiently sample k -sparse signals.

Definition 3 ([72]). An $m \times n$ matrix Φ is said to satisfy RIP(p, k, δ) (or in short RIP(p)) if for any

k -sparse vector \mathbf{x} ,

$$\|\mathbf{x}\|_p(1 - \delta) \leq \|\Phi\mathbf{x}\|_p \leq \|\mathbf{x}\|_p. \quad (4.2)$$

Definition 4. A bipartite graph (with n variable nodes and m encoded nodes) is called (k, d, ϵ) -regular expander if each variable node has exactly d neighbors in the encoded side and every subset S of variable nodes with at most k members has at least $(1 - \epsilon)d|S|$ neighbors [72], where $|\cdot|$ represents the cardinality of a set.

In the following, we investigate to see if $\Phi_N = [\frac{\Phi^1}{d_1} | \dots | \frac{\Phi^r}{d_r}]$ satisfies RIP property for signals with non-uniform sparsity (see Definition 5), where d_l is the degree of variable nodes in the subgraph corresponding to Φ^l (i.e, the column weight of Φ^l).

Definition 5. An $m \times n$ matrix Φ is said to satisfy $\text{RIP}(p, \kappa, \alpha, \delta)$ property if any (α, κ) -sparse vector $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r]^T$ satisfies (4.2), where $\alpha = [\alpha_1, \dots, \alpha_r]$, $\sum_{l=1}^r \alpha_l = 1$, and $\kappa = [k_1, \dots, k_r]$.

Definition 6. Consider a bipartite graph $G(X, Y, E)$ with $|X| = n$ and $|Y| = m$. Let us partition the n variable nodes in X into r sets X_1, X_2, \dots, X_r of sizes $\alpha_1 n, \alpha_2 n, \dots, \alpha_r n$ such that $\sum_{l=1}^r \alpha_l = 1$. Let each variable node in X_l have exactly d_l neighbors in Y for $l = 1, 2, \dots, r$. Such a bipartite graph is called a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph ($\kappa = [k_1, \dots, k_r]$, $\mathbf{d} = [d_1, \dots, d_r]$, $\alpha = [\alpha_1, \dots, \alpha_r]$), if every subset S of variable nodes, with at most k_l members chosen from X_l has at least $(1 - \epsilon)D$ neighbors. Here, D is the number of edges in the subgraph induced by S .

Remark 2. From Definition 6, it can be deduced that Φ^l (i.e., the section l of matrix Φ_N) is the adjacency matrix of a (k_l, d_l, ϵ) -regular expander graph for $l = 1, 2, \dots, r$.

Theorem 1. Let $\Phi_N = [\frac{\Phi^1}{d_1} | \dots | \frac{\Phi^r}{d_r}]$ be the adjacency matrix of a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander bipartite graph² $G(X, Y, E)$ with $|X| = n$ and $|Y| = m$. Then, Φ_N has $\text{RIP}(1, \kappa, \alpha, \delta)$ for $\delta = 2\epsilon$.

²The condition under which this is held will be discussed in Lemma 1 and Lemma 2.

Proof. This theorem is a generalization to the RIP property of the uniformly sparse expanders given in [72, 79]. The proof generalizes the proof in [79, Appendix D] to the $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander bipartite graphs. Let \mathbf{x} be an (α, κ) -sparse signal. The support T of \mathbf{x} has the indices of non-zero entries in \mathbf{x} . Let $\Phi^{(T)}$ represent the columns of matrix Φ that are chosen by T . Similarly, $\mathbf{x}^{(T)}$ is a set of entries in \mathbf{x} chosen by T . Therefore, we have $\|\mathbf{x}\|_1 = \|\mathbf{x}^{(T)}\|_1$ and $\Phi_N \mathbf{x} = \Phi_N^{(T)} \mathbf{x}^{(T)}$. On the other hand, given the matrix Φ_N is the adjacency matrix of a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander bipartite graph, a subset $X^{(T)}$ of variable nodes, chosen by the support set of \mathbf{x} , with at most k_l members chosen from X_l , has at least $(1 - \epsilon)D^{(T)}$ neighbors, where $D^{(T)}$ is the number of edges in the subgraph that is induced by $X^{(T)}$. Therefore, according to the Hall's Matching Theorem, there exists a partial matching M such that every variable node in $X^{(T)} \cap X_l$ is matched to at least $(1 - \epsilon)d_l$ unique neighbors in Y for $l = 1, 2, \dots, r$. Based on this matching, we decompose $\Phi_N^{(T)}$ as

$$\Phi_N^{(T)} = \Phi_M + \Phi_R, \quad (4.3)$$

where Φ_M only has the edges that participate in the matching M (see Figure 4.5 for a toy example). The matrix Φ_M is partitioned into r sections (the columns of Φ_M in section l correspond to the columns in Φ^l) and every column of Φ_M in section l has $d_l(1 - \epsilon)$ non-zeros for $l = 1, 2, \dots, r$ (Remark 2). Φ_R has the remaining non-zero entries of $\Phi_N^{(T)}$ that do not participate in matching M . From (4.3) and the triangle inequality, we have

$$\|\Phi_N^{(T)} \mathbf{x}^{(T)}\|_1 \geq \|\Phi_M \mathbf{x}^{(T)}\|_1 - \|\Phi_R \mathbf{x}^{(T)}\|_1 \quad (4.4)$$

Since every row of Φ_M has at most one non-zero entry and every column in section l of Φ_M has $d_l(1 - \epsilon)$ non-zeros, it is easy to see that an entry of $\mathbf{x}^{(T)}$ that corresponds to section l appears $d_l(1 - \epsilon)$ times in the $\|\Phi_M \mathbf{x}^{(T)}\|_1$. Therefore,

$$\|\Phi_M \mathbf{x}^{(T)}\|_1 \geq (1 - \epsilon)\|\mathbf{x}^{(T)}\|_1 \quad (4.5)$$

Similarly, Φ_R is composed of r sections and each column of Φ_R in section l has ϵd_l non-zeros for $l = 1, 2, \dots, r$. Hence, using the definition of ℓ_1 -norm, we have

$$\begin{aligned}\|\Phi_R \mathbf{x}^{(T)}\|_1 &= \sum_{i=1}^m \left| \sum_{j \in T} \Phi_{R_{i,j}} x_j \right| \leq \sum_{i=1}^m \sum_{j \in T} \Phi_{R_{i,j}} |x_j| \\ &= \sum_{j \in T} |x_j| \sum_{i=1}^m \Phi_{R_{i,j}} \leq \sum_{j \in T} |x_j| \epsilon \leq \epsilon \|\mathbf{x}^{(T)}\|_1.\end{aligned}\tag{4.6}$$

Substituting (4.5) and (4.6) in (4.4), we obtain

$$\|\Phi_N^{(T)} \mathbf{x}^{(T)}\|_1 \geq (1 - 2\epsilon) \|\mathbf{x}^{(T)}\|_1.\tag{4.7}$$

In addition, we can easily verify that

$$\begin{aligned}\|\Phi_N^{(T)} \mathbf{x}^{(T)}\|_1 &= \sum_{i=1}^m \left| \sum_{j \in T} \Phi_{N_{i,j}} x_j \right| \leq \sum_{i=1}^m \sum_{j \in T} \Phi_{N_{i,j}} |x_j| \\ &= \sum_{j \in T} |x_j| \sum_{i=1}^m \Phi_{N_{i,j}} = \|\mathbf{x}^{(T)}\|_1.\end{aligned}\tag{4.8}$$

Combining (4.7) and (4.8) and substituting $\|\mathbf{x}^{(T)}\|_1$ and $\Phi_N^{(T)} \mathbf{x}^{(T)}$ with $\|\mathbf{x}\|_1$ and $\Phi_N \mathbf{x}$, respectively, completes the proof. \square

The bipartite graph and its adjacency matrix are represented for a toy example in Figure 4.5. In this example, the parameters of the bipartite graph are $n = 10$, $m = 8$, $r = 2$, $\alpha_1 = \alpha_2 = 0.5$, $k_1 = k_2 = 2$, $d_1 = 4$, $d_2 = 2$ and $\epsilon = 1/2$.

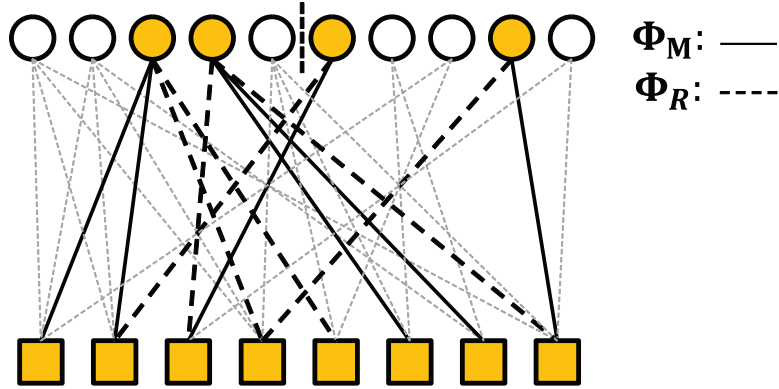


Figure 4.5: The bipartite graph representation of Φ_N for a toy example with $n = 10, m = 8, r = 2$, $\alpha_1 = \alpha_2 = 0.5, k_1 = k_2 = 2, d_1 = 4$, and $d_2 = 2$.

In Figure 4.5, the variables nodes corresponding to the support T are depicted with color. The edges that are induced by $\mathbf{x}^{(T)}$ (corresponding to $\Phi_N^{(T)}$) are shown in black color (both solid and dashed lines). Therefore,

$$\Phi_N = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

$$\Phi_N^{(T)} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}}_{\Phi_M} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \end{bmatrix}}_{\Phi_R}.$$

Accordingly $\Phi_N^{(T)}$ is written as sum of Φ_M and Φ_R as shown above. In Figure 4.5, the edges corresponding to Φ_M and Φ_R are represented by solid and dashed black lines, respectively.

In the following, we provide Lemma 1 that certifies the conditions in Theorem 1 are satisfied asymptotically (i.e., $k, m, n \rightarrow \infty$) with probability approaching 1 given the ratio of k/m is bounded by a given value. Moreover, Lemma 2, for the special case of $r = 2$, provides the sufficient condition on d_l 's in order to exist a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph for non-asymptomatic scenarios.

Lemma 1. *Consider a bipartite graph $G(X, Y, E)$ with $|X| = n$ and $|Y| = m$. Assume the n variable nodes in X are partitioned into r sets X_1, X_2, \dots, X_r of sizes $\alpha_1 n, \alpha_2 n, \dots, \alpha_r n$ such that $\sum_{l=1}^r \alpha_l = 1$. Let each variable node in X_l have exactly d_l neighbors in Y for $l = 1, 2, \dots, r$. Given $\sum_{l=1}^r k_l = k \leq \frac{n}{2}$, $0 < \epsilon < 1/2$, and $\frac{k}{m} < \rho_b(k, m, n; \mathbf{d}, \epsilon)$, then as $k, m, n \rightarrow \infty$, we have*

$$\text{Prob}(G \text{ fails to be a } (\kappa, \mathbf{d}, \alpha, \epsilon)\text{-irregular expander}) \rightarrow 0 \quad (4.9)$$

where $\kappa = [k_1, \dots, k_r]$, $\mathbf{d} = [d_1, \dots, d_r]$, $\alpha = [\alpha_1, \dots, \alpha_r]$, and $\rho_b(k, m, n; \mathbf{d}, \epsilon)$ is the largest limiting value of $\frac{k}{m}$ for which

$$\Psi(S_\kappa, n, m, \epsilon) = 0. \quad (4.10)$$

$\Psi(S, n, m, \epsilon) \triangleq H\left(\frac{|S|}{n}\right) + \frac{D}{n} H(\epsilon) + \frac{\epsilon D}{n} \log\left(\frac{D}{m}\right)$, where $H(p) \triangleq -p \log_e(p) - (1-p) \log_e(1-p)$ is the Shannon entropy function, S is the subset of X , D represents the sum of the degrees over the nodes in S , and S_κ is the set of subset of X with exactly k_l entries chosen from X_l for $l = 1, 2, \dots, r$.

Proof. For the case of bipartite graphs with variable nodes' degree d , the proof is provided in [80, Corollary 2.2]. We extend the proof to include the irregular expanders graphs as well. Consider all subsets $S \subset X$, which are the union of subsets S_1, S_2, \dots, S_r with given fixed sizes $|S_1| = \sigma_1, |S_2| = \sigma_2, \dots, |S_r| = \sigma_r$ such that each subset $S_l \subset X_l$ can have upto k_l coefficients in X_l for

$l = 1, 2, \dots, r$. We refer to set of such S 's as S_σ , where $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_r]$. For a specific subset $S \in S_\sigma$, with $D_\sigma = \sum_{l=1}^r \sigma_l d_l$ representing the number of edges in the graph induced by S , given G does not expand on S_σ , then in a sequence of D_σ emanating edges from S , more than ϵD_σ of them are associated to the vertices from Y that are identical to a preceding values in that sequence (We refer to these vertices as *collision set*).

Therefore, the probability that an encoded node in Y , chosen uniformly at random, belongs to the collision set is at most $\frac{D_\sigma}{m}$. Considering the selection of each encoded node is independent then the probability that a set of ϵD_σ randomly chosen encoded nodes belongs to the collision set is $(\frac{D_\sigma}{m})^{\epsilon D_\sigma}$. There are $\binom{D_\sigma}{\epsilon D_\sigma}$ ways to choose ϵD_σ vertices from D_σ nodes and $\binom{\alpha_1 n}{\sigma_1} \binom{\alpha_2 n}{\sigma_2} \dots \binom{\alpha_r n}{\sigma_r}$ ways to choose a set S from X .

Hence, for all subsets $S \in S_\sigma$ with $\Sigma_\sigma = \sum_{l=1}^r \sigma_l$, the probability that at least one such subset has more than ϵD_σ neighbors in the collision set (i.e., fails to expand on S_σ), is upper-bounded (using a union bound) as follows,

$$\text{Prob}(G \text{ fails to expand on } S_\sigma) \leq \binom{\alpha_1 n}{\sigma_1} \binom{\alpha_2 n}{\sigma_2} \dots \binom{\alpha_r n}{\sigma_r} \binom{D_\sigma}{\epsilon D_\sigma} \left(\frac{D_\sigma}{m}\right)^{\epsilon D_\sigma}. \quad (4.11)$$

Using Chu-Vandermonde identity, it is straightforward to show that $\binom{\alpha_1 n}{\sigma_1} \binom{\alpha_2 n}{\sigma_2} \dots \binom{\alpha_r n}{\sigma_r} \leq \binom{n}{D_\sigma}$.

Therefore,

$$\text{Prob}(G \text{ fails to expand on } S_\sigma) \leq \binom{n}{D_\sigma} \binom{D_\sigma}{\epsilon D_\sigma} \left(\frac{D_\sigma}{m}\right)^{\epsilon D_\sigma}. \quad (4.12)$$

We know $\binom{n}{np} \leq \frac{5}{4} (2\pi p(1-p)n)^{-\frac{1}{2}} \exp(nH(p))$ [80, Lemma 2.10]. Applying this upper bound

to the righthand side of (4.12), we obtain

$$\begin{aligned}
\binom{n}{D_\sigma} \binom{D_\sigma}{\epsilon D_\sigma} \left(\frac{D_\sigma}{m}\right)^{\epsilon D_\sigma} &\leq \frac{5}{4} (2\pi \frac{\Sigma_\sigma}{n} (1 - \frac{\Sigma_\sigma}{n}) n)^{-\frac{1}{2}} \exp(nH(\frac{D_\sigma}{n})) \\
&\quad \times \frac{5}{4} (2\pi \epsilon (1 - \epsilon) D_\sigma)^{-\frac{1}{2}} \exp(\epsilon H(\epsilon)) \\
&\quad \times \exp(\epsilon D_\sigma \log(\frac{D_\sigma}{m})) \\
&= p_m(S_\sigma) \exp(n\Psi(S_\sigma, n, m, \epsilon)),
\end{aligned} \tag{4.13}$$

where $p_m(S_\sigma) = (\frac{5}{4})^2 (2\pi)^2 (\frac{D_\sigma}{n} (1 - \frac{D_\sigma}{n}) n \epsilon (1 - \epsilon) D_\sigma)^{-\frac{1}{2}}$ and $\Psi(S_\sigma, n, m, \epsilon) = H(\frac{D_\sigma}{n}) + \frac{D_\sigma}{n} H(\epsilon) + \frac{\epsilon D_\sigma}{n} \log(\frac{D_\sigma}{m})$. Considering all possible values of $0 \leq \sigma_l \leq k_l$ for $l = 1, 2, \dots, r$, we have

$$\begin{aligned}
&\text{Prob}(G \text{ fails to be a } (\boldsymbol{\kappa}, \mathbf{d}, \boldsymbol{\alpha}, \epsilon)\text{-irregular expander}) \\
&\leq \sum_{\sigma_1=0}^{k_1} \sum_{\sigma_2=0}^{k_2} \dots \sum_{\sigma_r=0}^{k_r} \text{Prob}(G \text{ fails to expand on } S_\sigma) \\
&\leq \sum_{\sigma_1=0}^{k_1} \sum_{\sigma_2=0}^{k_2} \dots \sum_{\sigma_r=0}^{k_r} \binom{n}{D_\sigma} \binom{D_\sigma}{\epsilon D_\sigma} \left(\frac{D_\sigma}{m}\right)^{\epsilon D_\sigma}.
\end{aligned} \tag{4.14}$$

Given $k \leq \frac{n}{2}$, the largest summand in the right hand side of (4.14) is associated with the subset S^* in which exactly k_l entries are chosen from X_l for $l = 1, 2, \dots, r$. Therefore,

$$\begin{aligned}
\sum_{\sigma_1=0}^{k_1} \sum_{\sigma_2=0}^{k_2} \dots \sum_{\sigma_r=0}^{k_r} \binom{n}{D_\sigma} \binom{D_\sigma}{\epsilon D_\sigma} \left(\frac{D_\sigma}{m}\right)^{\epsilon D_\sigma} &< \sum_{\sigma_1=0}^{k_1} \sum_{\sigma_2=0}^{k_2} \dots \sum_{\sigma_r=0}^{k_r} \binom{n}{D_\kappa} \binom{D_\kappa}{\epsilon D_\kappa} \left(\frac{D_\kappa}{m}\right)^{\epsilon D_\kappa} \\
&< \prod_{l=1}^r (k_l + 1) p_m(S_\kappa) \exp(n\Psi(S_\kappa, n, m, \epsilon)).
\end{aligned} \tag{4.15}$$

The exponential term in (4.15) will be dominant as $k, m, n \rightarrow \infty$. Thus, having $\Psi(S_\kappa, n, m, \epsilon) < 0$ yields to $\text{Prob}(G \text{ fails to be a } (\boldsymbol{\kappa}, \mathbf{d}, \boldsymbol{\alpha}, \epsilon)\text{-irregular expander}) \rightarrow 0$. In [80, Corollary 2.11], it has been shown that $\Psi(S, n, m, \epsilon)$ is a monotonically increasing function of k/m . Therefore, given

$\Psi(S_{\kappa}, m, n, \epsilon) = 0$ for $k/m = \rho_b$, we conclude that $\Psi(S_{\kappa}, n, m, \epsilon) < 0$ for any value of $k/m < \rho_b$ as $k, m, n \rightarrow \infty$. Accordingly, the probability that graph G fails to be a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander goes to 0. \square

For the special case of $r = 2$, we define two types of signal coefficients. The *more important coefficients* (MICs) and *less important coefficients* (LICs) with lengths $n_1 = \alpha n$ and $n_2 = (1 - \alpha)n$, respectively ($\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$). In the measurement matrix $\Phi_N = [\frac{\Phi^1}{d_M}, \frac{\Phi^2}{d_L}]$, Φ^1 and Φ^2 have exactly d_M and d_L ones per column, respectively. The following lemma gives the sufficient conditions on the values of d_M and d_L that ensures the existence of a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander.

Lemma 2. *Consider bipartite graphs $G(X, Y, E)$ with $|X| = n$ and $|Y| = m$. Assume the n variable nodes in X are partitioned into two sets X_1 and X_2 of sizes αn and $(1 - \alpha)n$. Let each variable node in X_1 and X_2 have exactly d_M and d_L neighbors, respectively in Y . For all d_M and d_L that satisfy (4.16), there exists a bipartite graph that is $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander.*

$$\left\{ \begin{array}{l} (a) : \log \left(k_1 k_2 \binom{\alpha n}{k_1} \binom{(1 - \alpha)n}{k_2} \right) + \log((1 - \epsilon)D_{\kappa} - \tilde{d}) + \epsilon D_{\kappa} \log \left(\frac{(1 - \epsilon)D_{\kappa}}{m} \right) \\ \quad + (1 - \epsilon)D_{\kappa} - \frac{1}{2} \log(2\pi(1 - \epsilon)D_{\kappa}) < 0; \\ (b) : m \geq 2(1 - \epsilon)D_{\kappa}, \end{array} \right. \quad (4.16)$$

where $D_{\kappa} = k_1 d_M + k_2 d_L$, $\tilde{d} = \max(d_M, d_L)$, $\kappa = [k_1, k_2]$, $\mathbf{d} = [d_1, d_2]$, $\alpha = [\alpha, 1 - \alpha]$, $k_1 \leq \frac{\alpha n}{2}$, and $k_2 \leq \frac{(1 - \alpha)n}{2}$.

Proof. The proof is related to the proof of [81, Lemma 1], which considers the case of regular expanders (i.e., $r = 1$). We extend that proof to include the case of irregular expanders with $r = 2$. For $G(X, Y, E)$ with $|X| = n$ and $|Y| = m$, there exist $\binom{m}{d_M}^{\alpha n} \binom{m}{d_L}^{(1 - \alpha)n}$ incidents of irregular bipartite graphs with αn degree d_M and $(1 - \alpha)n$ degree d_L variable nodes. Consider all subsets

$S \subset X$, which are the union of subsets S_1 and S_2 with given fixed sizes $|S_1| = \sigma_1$ and $|S_2| = \sigma_2$ such that each subset $S_l \subset X_l$ can have upto k_l coefficients in X_l for $l = 1, 2$. We refer to set of such S 's as S_σ and $D_\sigma = \sigma_1 d_M + \sigma_2 d_L$ represents the number of edges in the graph induced by S_σ . For a subset S_σ , there are no more than $\sum_{m_s=\tilde{d}}^{\lceil(1-\epsilon)D_\sigma\rceil-1} \binom{m}{m_s} \binom{m_s}{d_M}^{\sigma_1} \binom{m_s}{d_L}^{\sigma_2}$ ways that this subset has fewer than $(1-\epsilon)D_\sigma$ neighbors in Y . Therefore, the number of cases that the graph fails to be $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander is less than or equal to

$$\sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2} \binom{m}{d_M}^{\alpha n - \sigma_1} \binom{m}{d_L}^{((1-\alpha)n) - \sigma_2} \sum_{m_s=\tilde{d}}^{\lceil(1-\epsilon)D_\sigma\rceil-1} \binom{m}{m_s} \binom{m_s}{d_M}^{\sigma_1} \binom{m_s}{d_L}^{\sigma_2}.$$

Hence, if the upper-bound on the the total number of non-expander graphs is less than all possible incidents of irregular bipartite graphs, there exists a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph. This is represented in the following

$$\begin{aligned} & \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2} \binom{m}{d_M}^{\alpha n - \sigma_1} \binom{m}{d_L}^{((1-\alpha)n) - \sigma_2} \\ & \sum_{m_s=\tilde{d}}^{\lceil(1-\epsilon)D_\sigma\rceil-1} \binom{m}{m_s} \binom{m_s}{d_M}^{\sigma_1} \binom{m_s}{d_L}^{\sigma_2} < \binom{m}{d_M}^{\alpha n} \binom{m}{d_L}^{(1-\alpha)n}. \end{aligned} \quad (4.17)$$

We can rewrite (4.17) as follows,

$$Q_t \triangleq \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \frac{\binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2}}{\binom{m}{d_M}^{\sigma_1} \binom{m}{d_L}^{\sigma_2}} \sum_{m_s=\tilde{d}}^{\lceil(1-\epsilon)D_\sigma\rceil-1} \binom{m}{m_s} \binom{m_s}{d_M}^{\sigma_1} \binom{m_s}{d_L}^{\sigma_2} < 1, \quad (4.18)$$

Note that $\binom{m_s}{d_M}$ and $\binom{m_s}{d_L}$ are increasing functions of m_s and $\lceil(1-\epsilon)D_\sigma\rceil - 1 \leq (1-\epsilon)D_\sigma$. Moreover, given $m \geq 2\lceil(1-\epsilon)D_\sigma\rceil$, for $d_M \leq m_s \leq (1-\epsilon)D_\sigma - 1$, $\binom{m}{m_s}$ is an increasing function

of m_s . Therefore,

$$\begin{aligned}
Q_t &\leq \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \frac{\binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2}}{\binom{m}{d_M}^{\sigma_1} \binom{m}{d_L}^{\sigma_2}} ((1-\epsilon)D_\sigma - \tilde{d}) \binom{m}{(1-\epsilon)D_\sigma} \binom{(1-\epsilon)D_\sigma}{d_M}^{\sigma_1} \binom{(1-\epsilon)D_\sigma}{d_L}^{\sigma_2} \\
&= \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2} ((1-\epsilon)D_\sigma - \tilde{d}) \binom{m}{(1-\epsilon)D_\sigma} \left(\frac{\binom{(1-\epsilon)D_\sigma}{d_M}}{\binom{m}{d_M}} \right)^{\sigma_1} \left(\frac{\binom{(1-\epsilon)D_\sigma}{d_L}}{\binom{m}{d_L}} \right)^{\sigma_2}.
\end{aligned} \tag{4.19}$$

In (4.19), we can substitute $\frac{\binom{(1-\epsilon)D_\sigma}{d_M}}{\binom{m}{d_M}}$ and $\frac{\binom{(1-\epsilon)D_\sigma}{d_L}}{\binom{m}{d_L}}$ with $\frac{\binom{m-d_M}{m-(1-\epsilon)D_\sigma}}{\binom{m}{m-(1-\epsilon)D_\sigma}}$ and $\frac{\binom{m-d_L}{m-(1-\epsilon)D_\sigma}}{\binom{m}{m-(1-\epsilon)D_\sigma}}$, respectively, because $\frac{\binom{m-h}{d}}{\binom{m}{d}} = \frac{\binom{m-d}{h}}{\binom{m}{h}} < \left(\frac{m-h}{m}\right)^d$. Hence (4.19) can be written as,

$$\begin{aligned}
Q_t &\leq \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2} ((1-\epsilon)D_\sigma - \tilde{d}) \\
&\quad \binom{m}{(1-\epsilon)D_\sigma} \left(\frac{\binom{m-d_M}{m-(1-\epsilon)D_\sigma}}{\binom{m}{m-(1-\epsilon)D_\sigma}} \right)^{\sigma_1} \left(\frac{\binom{m-d_L}{m-(1-\epsilon)D_\sigma}}{\binom{m}{m-(1-\epsilon)D_\sigma}} \right)^{\sigma_2} \\
&\leq \sum_{\sigma_1=1}^{k_1} \sum_{\sigma_2=1}^{k_2} \binom{\alpha n}{\sigma_1} \binom{(1-\alpha)n}{\sigma_2} ((1-\epsilon)D_\sigma - \tilde{d}) \\
&\quad \binom{m}{(1-\epsilon)D_\sigma} \left(\frac{(1-\epsilon)D_\sigma}{m} \right)^{\sigma_1 d_M + \sigma_2 d_L}.
\end{aligned} \tag{4.20}$$

Given $\alpha n \geq 2k_1$ and $(1-\alpha)n \geq 2k_2$ all summands in the right hand side of (4.20) are increasing functions of σ_1 and σ_2 . Therefore, we have

$$Q_t \leq k_1 k_2 ((1-\epsilon)D_\kappa - d_M) \binom{\alpha n}{k_1} \binom{(1-\alpha)n}{k_2} \binom{m}{(1-\epsilon)D_\kappa} \left(\frac{(1-\epsilon)D_\kappa}{m} \right)^{D_\kappa}. \tag{4.21}$$

By taking natural log from both sides of (4.21) and noting that $\binom{n}{k} < \frac{1}{\sqrt{2\pi k}} \left(\frac{n}{k}\right)^k \exp(k)$ (can be

shown using Stirling's formula), we have

$$\begin{aligned}
\log(Q_t) &\leq \log \left(k_1 k_2 \binom{\alpha n}{k_1} \binom{(1-\alpha)n}{k_2} \right) + \log \left((1-\epsilon)D_\kappa - \tilde{d} \right) \\
&\quad + D_\kappa \log \left(\frac{(1-\epsilon)D_\kappa}{m} \right) + \log \left(\binom{m}{(1-\epsilon)D_\kappa} \right) \\
&\leq \log \left(k_1 k_2 \binom{\alpha n}{k_1} \binom{(1-\alpha)n}{k_2} \right) + \log \left((1-\epsilon)D_\kappa - \tilde{d} \right) \\
&\quad + D_\kappa \log \left(\frac{(1-\epsilon)D_\kappa}{m} \right) - (1-\epsilon)D_\kappa \log \left(\frac{(1-\epsilon)D_\kappa}{m} \right) \\
&\quad + (1-\epsilon)D_\kappa - \frac{1}{2} \log(2\pi(1-\epsilon)D_\kappa).
\end{aligned} \tag{4.22}$$

In (4.22), we provided an upper bound for the natural logarithm of Q_t . If this upper bound is less than zero it means Q_t was less than 1 and there exists a $(\kappa, d, \alpha, \epsilon)$ -irregular expander graph. Applying this condition on (4.22), the inequality in (4.16) follows. \square

In Figure 4.6, we have depicted the conditions given in Lemma 2 on d_M - d_L plane for a numerical example with $n = 5 \times 10^5$, $m = 1.2 \times 10^5$, $k_1 = 80$, $k_2 = 20$, $\alpha = 0.5$ and different values of ϵ . The areas in d_M - d_L plane where (4.16) holds are shown with the white color.

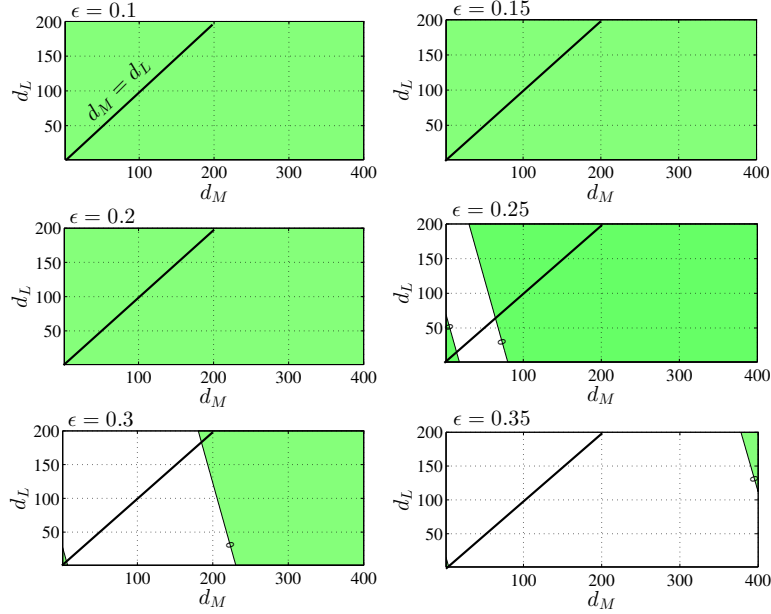


Figure 4.6: The areas in d_M - d_L plane where (4.16) holds are shown with white color. The design of NCS imposes $d_M \geq d_L$. Therefore the desirable values for d_M and d_L are below the $d_M = d_L$ line. For all d_M and d_L values in those areas, there exists a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph.

Please note that the values of d_M and d_L obtained using Lemma 2 are only sufficient for existence of a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph. Hence, it is possible to find the set of parameters that does not satisfy the conditions in Lemma 2, yet a $(\kappa, \mathbf{d}, \alpha, \epsilon)$ -irregular expander graph exists for those parameters.

Application of NCS in Burst Image Capture

As we mentioned earlier, one of the interesting applications of NCS is in burst image capture. In capturing images with a high frame rate, the changes from one frame to the next is very small and this makes the difference frame very sparse [82]. This property has been employed in high-frame-

rate image and video sampling [83]. However, this sampling strategy is oblivious to the ROIs in the image. We can apply the proposed NCS to the sampling of bursty images as shown in Figure 4.1. Since the pixels are divided into ROI and non-ROI pixels, the NCS will have only two importance levels (i.e., more important coefficients and less important coefficients). Accordingly the first frame is sampled and recovered using a typical CS scheme. After the first frame is recovered the ROI, which is determined based on the saliency of the pixels [84], is being extracted. This ROI is fed back to the sampler to set up the NCS' measurement matrix and the more important coefficients. Clearly, it is assumed that in the high frame rate image capture, the ROI does not change drastically from one frame to the next.

The Structure of the NCS-Based Image Sensor

An efficient image sensor can be implemented using the proposed NCS sampling. In high-frame-rate image capture, the saliency information of one recovered frame can be passed on to the next frame and determine the nonuniformity of the columns of Φ_N . The most salient pixels contribute to more measurements compared to less salient pixels. The structure of the proposed NCS-based image sampler is depicted in Figure 4.7. The columns that are denoted by Φ^1 and Φ^2 are used to sample the ROI and non-ROI pixels, respectively. The number of nonzero entries per column of Φ^1 and Φ^2 are d_1 and d_2 , respectively. The nonzero entries at each row of Φ_N determines the pixels that their analog reading is being applied to the input of the $\Sigma\Delta$ ADC. The output of ADC block in this case will be the quantized value of the average input to the ADC [69]. The output of each ADC will be treated as one CS measurement. It is worth noting that the parallel structure of ADCs and multiplexers that is proposed in Figure 4.7 implies that all m CS measurements can be captured in a single-shot.

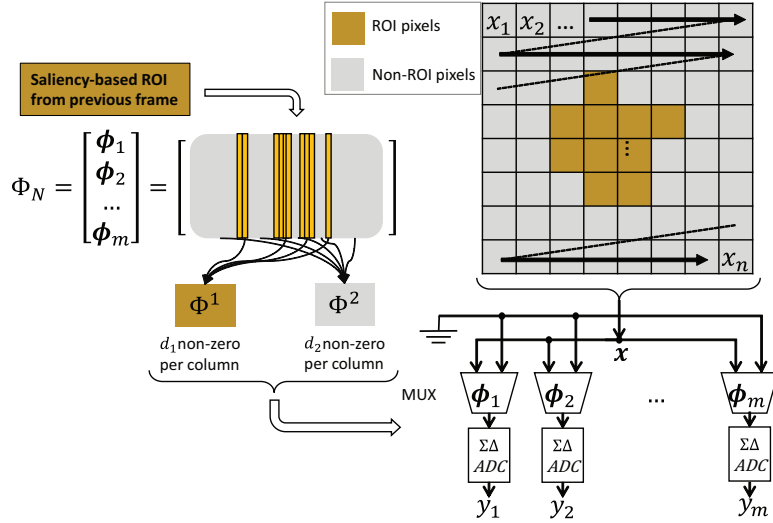


Figure 4.7: Block diagram representation of the proposed NCS-based image sampling. The rows of non-uniformly sparse Φ_N are used to determine the pixels that their values are multiplexed in analog domain and applied to the input of $\Sigma\Delta$ modulator

Performance Evaluation

In this section, we consider a special case of NCS in which the signal \mathbf{x} has two sections ($r = 2$). We compare the performance of NCS with the performance of uniform CS. In the uniform CS scenario, each column of the measurement matrix has exactly d non-zero entries (with value $\frac{1}{d}$). In the non-uniform scenario with $r = 2$, we refer to MICs and LICs by \mathbf{x}_{MIC} and \mathbf{x}_{LIC} and $\mathbf{x} = [\mathbf{x}_{MIC}^T, \mathbf{x}_{LIC}^T]^T$. We choose $d_M \geq d \geq d_L$ such that the average number of non-zero entries L in each row of the measurement matrix in both uniform and non-uniform measurement matrices remain equal. Therefore, the probability of having a non-zero entry in Φ^1 and Φ^2 becomes $p_M = \frac{d_M}{m} \geq \frac{d}{m}$ and $p_L = \frac{d_L}{m} \leq \frac{d}{m}$.

We divide p_M and p_L by $p = \frac{d}{m} = \frac{L}{n}$ and introduce two new parameters $q_M = \frac{p_M}{p} = \frac{d_M}{d}$ and $q_L = \frac{p_L}{p} = \frac{d_L}{d}$. Here, q_M and q_L solely represent the non-uniformity of the measurement matrix regardless of its sparsity. Clearly, $0 < q_L \leq 1$, and $q_M = \frac{1-(1-\alpha)q_L}{\alpha}$.

We consider two types of sparse signals in our simulations: signals with *uniform sparsity*³ and signals with *non-uniform sparsity*. A signal with uniform sparsity has k non-zeros uniformly distributed in its n coefficients. On the other hand, a signal with non-uniform sparsity has k_1 and k_2 non-zeros among n_1 MICs and n_2 LICs, respectively, with $\frac{k_1}{n_1} \neq \frac{k_2}{n_2}$. In our simulations, we set $n = 1000$, $\alpha = 0.15$ (i.e., $n_1 = 150$, and $n_2 = 850$), $k = 100$, $k_1 = 30$, $k_2 = 70$, and $d = 0.04m$.

We perform the CS sampling employing Φ_N and perform the recovery employing the CVX optimization tool [85, 86] to obtain $\hat{\mathbf{x}}$, which is the estimate of the signal \mathbf{x} . We plot the *normalized recovery error* (NRE) of MICs and LICs given by $\frac{\|\mathbf{x}_{MIC} - \hat{\mathbf{x}}_{MIC}\|_2}{\|\mathbf{x}_{MIC}\|_2}$ and $\frac{\|\mathbf{x}_{LIC} - \hat{\mathbf{x}}_{LIC}\|_2}{\|\mathbf{x}_{LIC}\|_2}$, respectively, versus different parameters, where $\hat{\mathbf{x}}_{MIC}$ and $\hat{\mathbf{x}}_{LIC}$ correspond to the MICs and LICs in the recovered signal $\hat{\mathbf{x}}$, respectively. We compare these results with the overall NRE of the signal ($\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2}$) using uniform CS measurement matrix, for which $q_M = 1$. In our numerical simulations, when the parameters dictate a non-integer value for d_L , it has been rounded to nearest integer value.

In the first simulation, we fix the number of measurements to $m = 300$ ($d = 12$) and plot NRE of MICs and LICs versus q_M in Figures 4.8 and 4.9 for signals with uniform and non-uniform sparsities, respectively. In Figures 4.8 and 4.9, we have also depicted the overall NRE of signal \mathbf{x} (the solid black line with + markers).

³As an interesting special case, NCS can be applied to *uniformly sparse* signals to achieve non-uniform *recovery*. By applying NCS to a uniformly sparse signal, different parts of that signal are recovered with different accuracy levels.

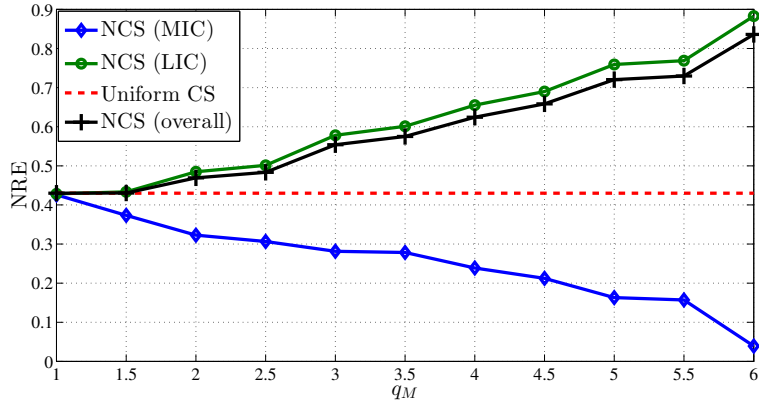


Figure 4.8: NRE of MICs and LICs versus q_M for *uniformly sparse* signals with $n = 1000$ and $k = 100$. For Φ_N , we set $\alpha = 0.15$, $d = 12$, $m = 300$, and $1 \leq q_M \leq 6$. Note that $q_M = 1$ corresponds to the NRE of uniform CS.

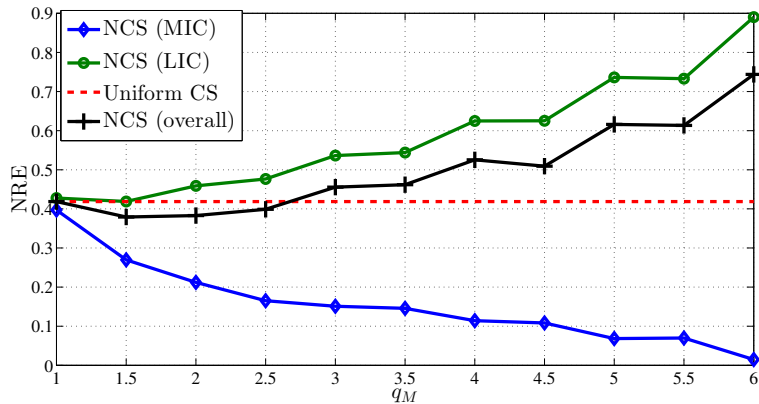


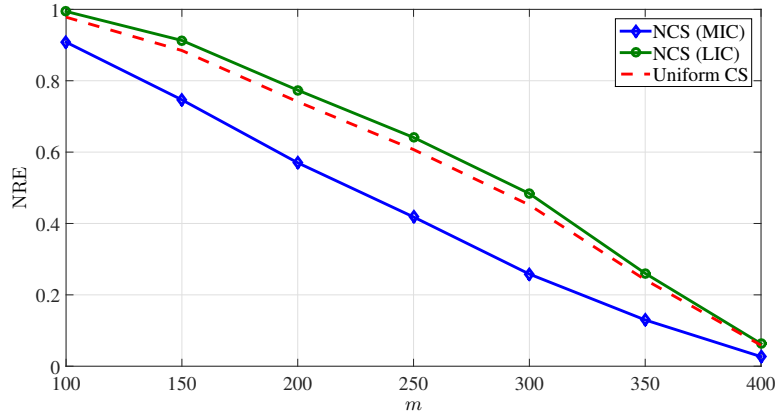
Figure 4.9: NRE of MICs and LICs versus q_M , for *non-uniformly sparse* signals with $n = 1000$, $k_1 = 30$, $k_2 = 70$, and $\alpha = 0.15$. For Φ_N we set $\alpha = 0.15$, $d = 12$, $m = 300$, and $1 \leq q_M \leq 6$.

In Figure 4.8, the significant coefficients of \mathbf{x} are distributed uniformly and the goal is to recover the first part of the signal (MICs) with a higher accuracy. Figure 4.8 shows that as q_M increases, MICs are recovered with a smaller error at the cost of LICs increasing recovery error. Therefore,

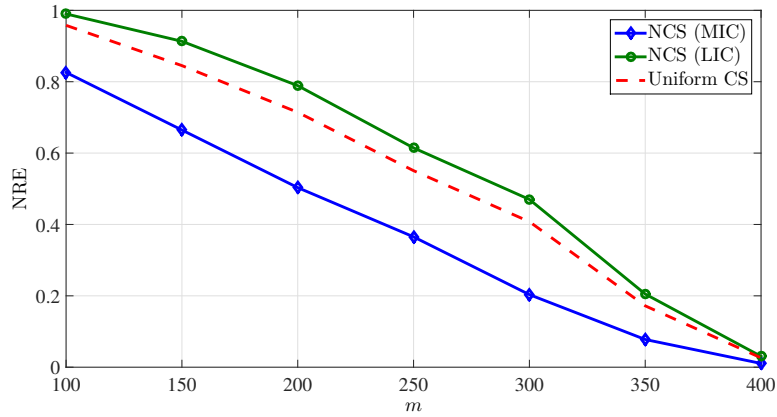
the desired non-uniform recovery has been provided for MICs and LICs. As an example, we observe that at $q_M = 2$, MICs are recovered with 25% higher accuracy compared to uniform CS ($q_M = 1$), while the performance loss of LICs is less than 15%. Moreover at $q_M = 1.5$, MICs are recovered with 15% smaller error without losing any performance for LICs.

In Figure 4.9, the significant coefficients of \mathbf{x} are non-uniformly distributed (30% of the significant coefficients are positioned in the initial 15% of the signal coefficients). The desired more accurate recovery of MICs is provided using the proposed method. At $q_M = 2$, MICs are recovered with 50% higher accuracy compared to uniform CS ($q_M = 1$), while the performance loss of LICs is only 5%. A very interesting results can be seen in Figure 4.9. For $1 < q_M < 2.5$, not only the recovery performance of MICs are significantly improved, but also the overall performance is better than the uniform sampling ($q_M = 1$). This is because in signals with non-uniform sparsity, MICs are not only sampled more frequently but also have a higher concentration of non-zeros. Therefore, implementing NCS improves the overall performance as well.

Next, we fix $q_M = 2.5$ and plot NREs of MICs and LICs versus the number of measurements m for signals with uniform and non-uniform sparsities as illustrated in Figures 4.10(a) and 4.10(b), respectively. We have also included NRE of the uniform CS ($q_M = 1$) in both figures. We see a greater improvement in NRE of MICs when the signal has non-uniform sparsity compared to the signals with uniform sparsity. For example, at $m = 250$ MICs' performance improves over uniform CS by 40% and 34% for signals with non-uniform and uniform sparsities, respectively.



(a) NRE versus m for NCS with uniformly sparse signals ($k = 100$).

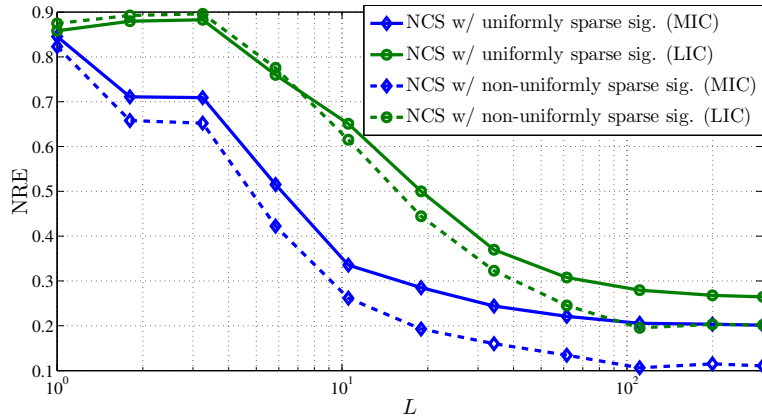


(b) NRE versus m for NCS with non-uniformly sparse signals ($k_1 = 30$ and $k_2 = 70$).

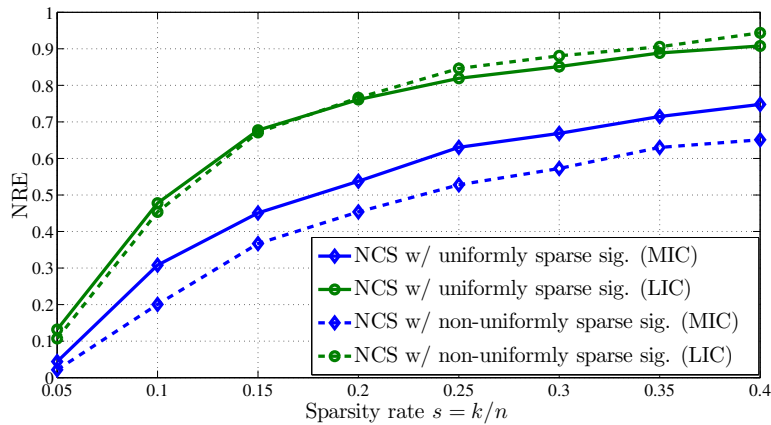
Figure 4.10: NRE versus m for $n = 1000$, $\alpha = 0.15$, and $d = 12$, $q_M = 2.5$. The dashed line corresponds to the NRE of uniform CS.

In Figure 4.11(a), we plot NREs versus the average row weight L (by adjusting d_M and d_L) when $q_M = 2$ and $m = 300$. We observe that for $L \geq 40$ NREs are almost insensitive to L and the recovery error remains almost constant. In Figure 4.11(b), we plot NREs versus the sparsity $s = \frac{k}{n}$, by varying k . We observe that NRE is an increasing function of s . This is expected

since increasing s results in a less compressible (less sparse) signal. However, we observe that non-uniform recovery is still effectively provided for MICs and LICs compared to uniform CS.



(a) NRE versus average degree of measurements, L .



(b) NRE versus the sparsity, $s = \frac{k}{n}$.

Figure 4.11: NRE versus L and $s = \frac{k}{n}$ for MICs and LICs of signals with uniform (solid lines) and non-uniform (dashed lines) sparsities with $n = 1000$, $m = 300$, $\alpha = 0.15$, and $q_M = 2.5$.

Figure 4.12 shows a visual comparison between the performances of NCS and uniform CS. In addition, we compare our scheme with saliency-based compressive sampling scheme in [75], which

divides the pixels into blocks and assigns measurements to each block based on the number salient pixels in that block. In Figure 4.12(a) a 128×128 Lenna image is represented. The image signal \mathbf{x} has a sparse representation in some basis \mathcal{B} (such as Wavelet or DCT). In the case of NCS, the random measurements are generated by $\mathbf{y} = \Phi_N \mathbf{x}$. The pixels corresponding to the face area are considered as MICs, which are located inside the rectangle shown on the image. The number of MICs are 1632 that makes $\alpha \simeq 0.1$. In Both Figures 4.12(b) and 4.12(c), we generated $m = 6000$ measurements employing uniform CS and NCS, respectively, and used the standard ℓ_1 -norm minimization to recover the sparse coefficients of the image. In the case of uniform CS in Figure 4.12(b), the NRE of the entire recovered image is 0.1249. However, as shown in Figure 4.12(c), by employing NCS (with $q_M = 9$), NRE of recovered MICs is dropped significantly to 0.0253 and the NRE of LICs is slightly increased to 0.1347 ($\approx 8\%$ increase). In addition, we provide the ROI in Figure 4.12(a) to the saliency-based compressive sampling algorithm [75]. This algorithm divides the image into 16 32×32 blocks. Each block is measured separately and the number of measurements per block is determined based on the number of salient pixels that falls into that block. Using this algorithm, NRE of recovered MICs is dropped to 0.0410 while the NRE of the entire image is 0.1304 (see Figure 4.12(d)). As it can be seen the NRE on MICs in saliency-based block CS is also improved but not as much the NCS.

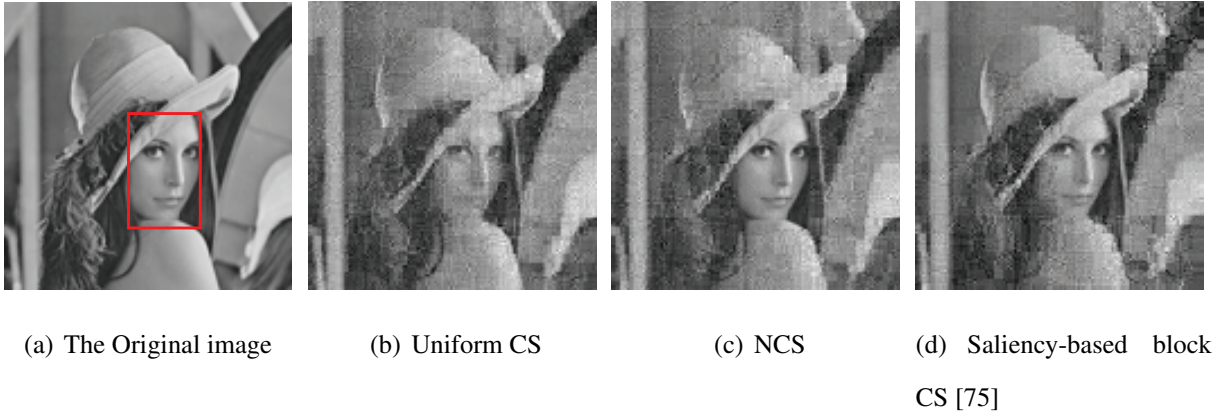


Figure 4.12: A pictorial example for comparing the performances of NCS and uniform CS. Here the image is 128×128 pixels and the number of measurements is $m = 6000$. For NCS, we have $\alpha \simeq 0.1$ and $q_M = 9$. Substantial improvement in the recovery performance of the region of interest (face) is achieved with a slight degradation in the recovery performance of the other parts.

In the next simulation, we compare NCS with the weighted ℓ_1 scheme [1]. We consider the following parameters $n = 200$, $m = 100$, $\alpha = 0.5$, and $k_2 = 10$ (as given in [1]). Figure 4.13 compares the *probability of recovery* for the two scenarios. The probability of recovery is defined as percentage of the trials that the normalized recovery error is less than 0.001. The x -axis represents k_1 , the number of non-zeros in the first section of the sparse signal. In Figure 4.13, for each value of k_1 the recovery probabilities are chosen independently with empirically optimized weights (in the case of weighted ℓ_1) and optimized q_M values (in the case of NCS). Weighted ℓ_1 is a technique that exploits non-uniformity in the sparsity of the signal at the recovery phase, as opposed to NCS that utilizes such information in the sampling phase. As we see in Figure 4.13, weighted ℓ_1 slightly outperforms NCS. However, interestingly, weighted ℓ_1 can be combined with NCS to result in a better performance. As we can see the probability of recovery is maximized when the NCS is used in the sampling step and weighted ℓ_1 is used in the recovery step.

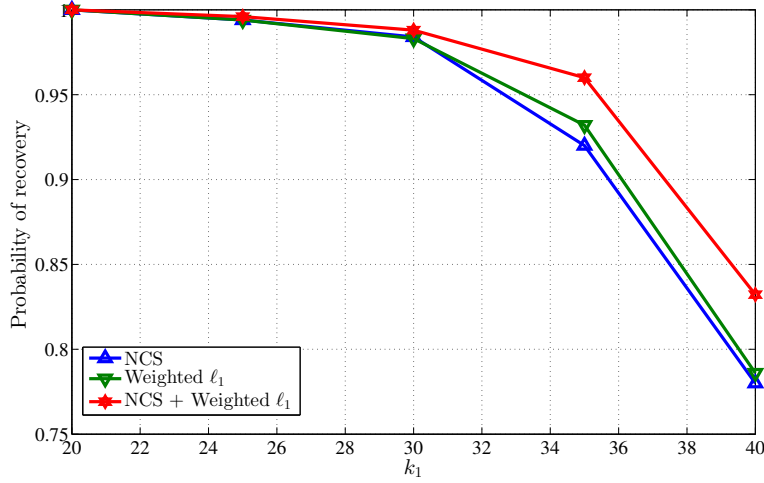
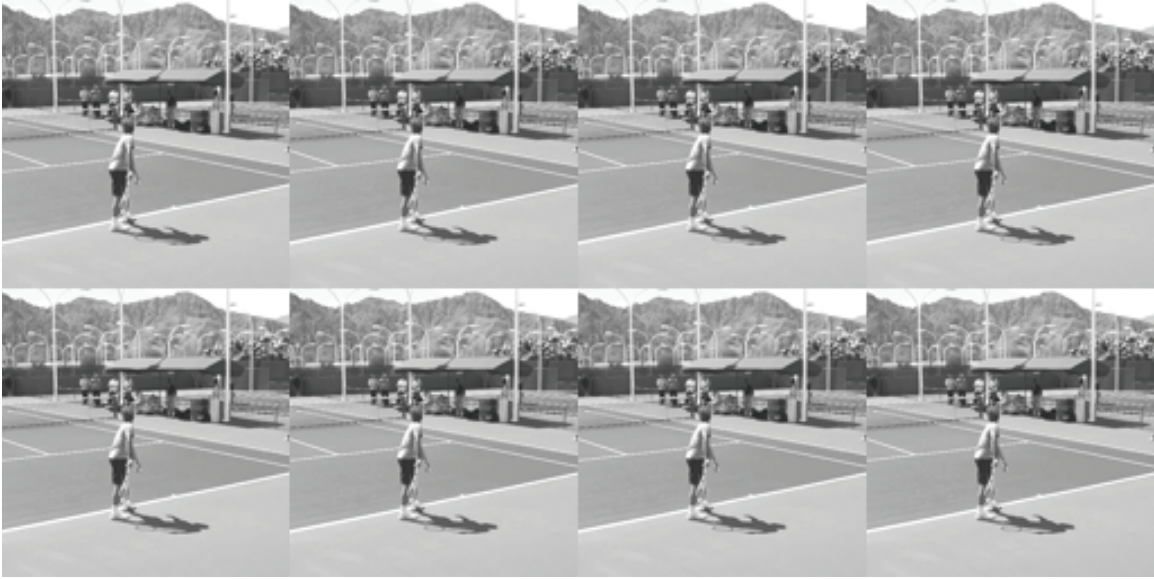
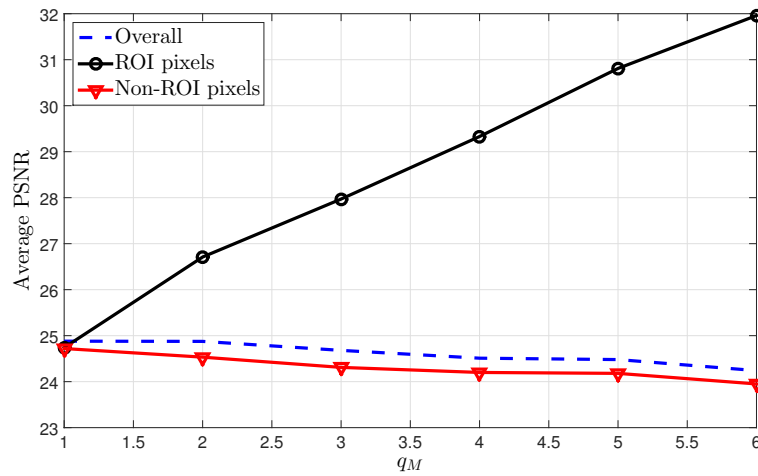


Figure 4.13: The probability of recovery versus k_1 for our proposed NCS and weighted ℓ_1 [1] for $n = 200$, $m = 100$, $\alpha = 0.5$, and $k_2 = 10$.

Next, the proposed NCS-based image sampling is employed to capture burst-mode images. For our simulations, we use a raw video with frame rate 500 fps. The frames of this video are used to simulate the readout of an image sensor. For the first frame, the saliency information is not available. Therefore, the CS sampling matrix is chosen to be uniformly sparse. After the samples of the first frame recovered the most salient pixels (i.e., MICs) will be determined and will be utilized to setup Φ_N for the next frame. Employing this process over 8 consecutive frames of the video (shown in Figure 4.14(a)) and using $m = 6000$ measurements to samples each frame, we depicted the average PSNR of the recovered frames and the PSNR of 10% most salient ROI pixels and non-ROI pixels separately versus q_M as shown in Figure 4.14(b).



(a) Sampled frames



(b) Average PSNR versus q_m

Figure 4.14: The average PSNR of 8 consecutive frames that are sampled using proposed NCS. Each frame has 128×128 pixels and is sampled by $m = 6000$ measurements.

As we can see the overall PSNR without using the NCS (i.e., $q_M = 1$) is 24.8 dB. However, by

employing the NCS-based sampling and setting $q_M = 5$, the PSNR of 10% most salient pixels is improved to 30.9 dB (more than 6 dB improvement), while the PSNR of remaining Non-ROI pixel dropped only by less than 1 dB.

In the next simulation, we apply our proposed NCS and saliency-based CS [75] to a 128×128 frame. The original frame and its 10% most salient pixels are shown in Figure 4.15(a). The number of measurements per frame is set to $m = 6500$. The goal is to have the overall PSNR of the recovered image greater than or equal to 25 dB. Accordingly we set q_M such that PSNR of the ROI pixels is maximized while the overall PSNR is at least 25 dB. The visual performance of both schemes is depicted in Figure 4.15. The proposed NCS scheme, while achieving the overall PSNR of 25 dB, reaches the ROI PSNR of 32.12 dB, which is significantly higher than that of the saliency-based CS (28.93 dB).



(a) Original frame and ROI (b) NCS-based sampling (ROI PSNR: 32.12) (c) Saliency-based CS [75] (ROI PSNR: 28.93)

Figure 4.15: Recovered image using NCS and saliency-based CS schemes. The overall PSNR for both schemes is set to be greater than or equal to 25.

Conclusion

In this chapter, we proposed *non-uniform compressive sensing* (NCS) for recovery of sparse signals with either non-uniform sparsity or non-uniform recovery requirement at different parts of the signal. To realize NCS, more important signal coefficients are captured more frequently by the measurements. In this way, NCS will provide a higher recovery accuracy for more important coefficients compared to less important coefficients, which is of interest in many applications. NCS is based on only modifying the sampling step of conventional CS schemes and it can be integrated with any CS recovery scheme, such as Basis Pursuit or weighted ℓ_1 . We analyzed the proposed NCS measurement matrix and showed that it satisfies the weak restricted isometry property. Next, we performed extensive numerical simulations and showed that by correctly setting NCS parameters, the desired non-uniform recovery will be achieved. As an interesting application, we have shown that NCS can be effectively applied to the burst mode image capture to provide non-uniform recovery performance for the ROI pixels.

CHAPTER 5: MODEL-BASED NON-UNIFORM COMPRESSED SAMPLING AND RECOVERY OF NATURAL IMAGES UTILIZING A WAVELET-DOMAIN UNIVERSAL HIDDEN MARKOV MODEL

In this chapter, a novel model-based compressive sampling (CS) technique for natural images is proposed. Our algorithm integrates a *universal hidden Markov tree (uHMT)* model, which captures the relation among the sparse wavelet coefficients of images, into both *sampling* and *recovery* steps of CS. At the *sampling* step, we employ the uHMT model to devise a nonuniformly-sparse measurement matrix Φ_{uHMT} . In contrast to the conventional CS sampling matrices such as dense Gaussian, Bernoulli or uniformly sparse matrices that are oblivious to the signal model and the correlation among the signal coefficients, the proposed Φ_{uHMT} is designed based on the signal model and samples the more important coefficients (wavelet coefficients at coarser scales) with a higher probability compared to the less important coefficients.

At the *recovery* step, we integrate the uHMT model into two state-of-the-art Bayesian CS recovery schemes. Our simulation results confirm the superiority of our proposed *HMT model-based non-uniform compressive sampling and recovery*, referred to as *uHMT-NCS*, over other model-based CS techniques that solely consider the signal model at the recovery step. Our work is distinguished from other model-based CS schemes in that we take a novel approach to simultaneously integrating the signal model into both CS *sampling* and *recovery* steps. We show that such integration greatly increases the performance of the CS recovery, which is equivalent to reducing the required number of samples for a given reconstruction quality.

Introduction

Today's multimedia-rich applications have dramatically increased the traffic flow in the communication networks. To deal with this overwhelmingly large amount of data and reduce the computational complexity, new compression techniques are on demand. In this regard, the emerging field of compressive sampling (CS) [22, 53] that has revolutionized the traditional concept of sensing and sampling established by the Nyquist sampling theorem has attracted a lot of attention. According to the CS theory, the signals that have a sparse representation over a proper basis can be recovered from a small set of linear measurements.

The compression efficiency of CS cannot compete with conventional codec such as JPEG2000 or MPEG4 when dealing with already acquired image or video signals with high resolution and quality [87]. However, CS is still desirable in applications in which sensing is expensive (such as MRI or infra-red imaging). The conventional CS algorithms merely exploit signal *sparsity* in their designs. Nevertheless, it has recently been shown that in addition to the sparsity, we can utilize the extra *knowledge* about the signal structure as a priori information in the CS *recovery* step to enhance the overall CS recovery performance compared to the conventional CS recovery algorithms [2–4, 70]. In a very recent work [88], Indyk and Razenshteyn proved that for the signals with tree-structured sparsity (such as natural images), the recovery is achievable with fewer measurements compared to the recovery of general sparse signals. In [70], the tree structure of wavelet coefficients is used to design a CS recovery algorithm based on weighted ℓ_1 minimization, called HMT-based IRWL1, for one dimensional piecewise smooth signals. In [4], the tree structure of wavelet coefficients is exploited to create a statistical model for the sparse coefficients that results in more accurate recovery. In [2], the authors have modified the novel approximate message passing algorithm [43] such that the tree structure of the wavelet coefficients is utilized in the recovery process.

In this chapter, we take one step further and show that by exploiting the signal model in the *sampling* step of the CS in addition to the recovery step, we can achieve greater gains. This chapter mainly focuses on natural images. It is well known that the signal coefficients of natural images in the wavelet domain are not only sparse but also are correlated by a tree structure as described in [36, 89]. Romberg et al. developed a *universal hidden Markov model (uHMT)* for normalized natural images in [89]. The uHMT model provides relatively accurate predictions about the wavelet coefficients. We will demonstrate how uHMT model can be integrated into the design of novel sampling and recovery schemes. In this chapter, \mathbf{x} corresponds to the pixels of an image, Ψ is the wavelet basis, and $\boldsymbol{\theta} = \Psi^T \mathbf{x}$ corresponds to the sparse wavelet coefficients of the image.

There have been some recent studies on unconventional measurement matrices, such as structured measurement matrices [50] or Toeplitz matrices [90]) to mimic the real-world applications in which the acquisition of samples by a random Gaussian or Bernoulli sampling matrices are infeasible [50] (e.g., multipath channel estimation [90]). However, only a few contributions deliberately modify the sampling matrix structure with the goal of improving the overall CS performance [91–94]. In [91], the authors have shown that separately measuring each scale of signals’ wavelet coefficients improves the recovery performance. In [92], the authors have exploited the visual importance of different areas of an image in the sampling step and have modified the block-CS algorithm (proposed in [32]) to generate CS measurements with different compression levels. In [93], the authors have proposed a structurally random structure for the measurement matrix which provides fast computation and low complexity compared to random measurement matrices. In [95], the authors proposed a model-based dictionary learning considering the tree structure of the wavelet coefficients. In [94], the measurement matrix is designed such that it is matched to the dictionary that is learned through training images. Such a match reduces the coherence between the measurement matrix and the dictionary and enhances the CS recovery performance.

To the best of our knowledge [91, 92] are the only studies that modify the sampling step based

on the sparse signal model (in the wavelet domain) to improve the performance. In our work, we take a novel and more effective approach to generate the *CS measurement matrix* to be applied to natural images utilizing the properties of multi-scale wavelet transform as side information.

Contribution of this chapter

In this chapter, we exploit the uHMT model of wavelet coefficients of natural images to modify *both CS sampling and recovery steps* and significantly enhance the performance. The contribution of this chapter is two-fold. First, using the uHMT model, we propose a *non-uniform CS measurement matrix* that generates CS measurements such that they include the significant coefficients *with a higher probability* compared to the non-significant coefficients. Second, we propose two model-based recovery algorithms that employ the uHMT model to improve the CS recovery performance even further. The novelty of this work is in proposing a uHMT-model-based *non-uniform CS sampling*, and in *simultaneously* employing the tree structure of wavelet coefficients at both CS sampling and CS recovery steps. Figure 5.1 depicts the block diagram of the sampling and the recovery steps in our proposed scheme, referred to as *uHMT-NCS*. Although this work mainly focuses on natural images, the ideas from this work can be extended for the design of new CS matrices when dealing with different signal models.

The Universal Hidden Markov Tree (uHMT) Model

In this section, we provide a brief introduction to uHMT model for wavelet coefficients of natural images that we will use later to develop our proposed schemes.

The multi-resolution wavelet decomposition of images has many applications in image processing (e.g., JPEG2000 standard). The wavelet-domain coefficients of an image exhibit both sparsity

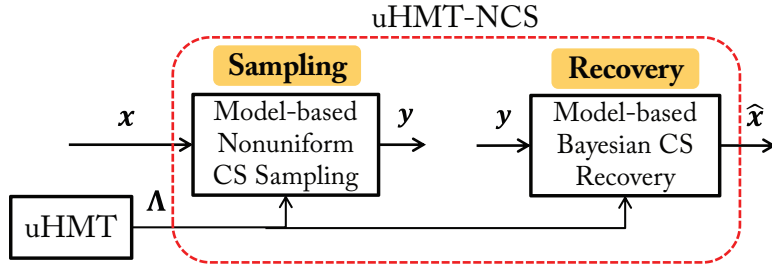


Figure 5.1: The block diagram of the proposed uHMT-NCS scheme. The vectors x , y , and \hat{x} correspond to the image signal, the CS measurements, and the recovered image, respectively. The uHMT model parameters (Λ) are utilized at both CS Sampling and CS recovery steps.

and a *tree structure* [36, 89]. Figure 5.2 shows a two-dimensional wavelet transform of an image with two scales. The top-left block represents the approximate coefficients (A). The next two top blocks (H_1 and H_2), the two blocks along the left side of the image (V_1 and V_2), and the two diagonal blocks (D_1 and D_2) represent the tree structures along the horizontal, the vertical, and the diagonal directions, respectively. As it can be seen in Figure 5.2, quad trees form in horizontal, vertical, and diagonal directions.

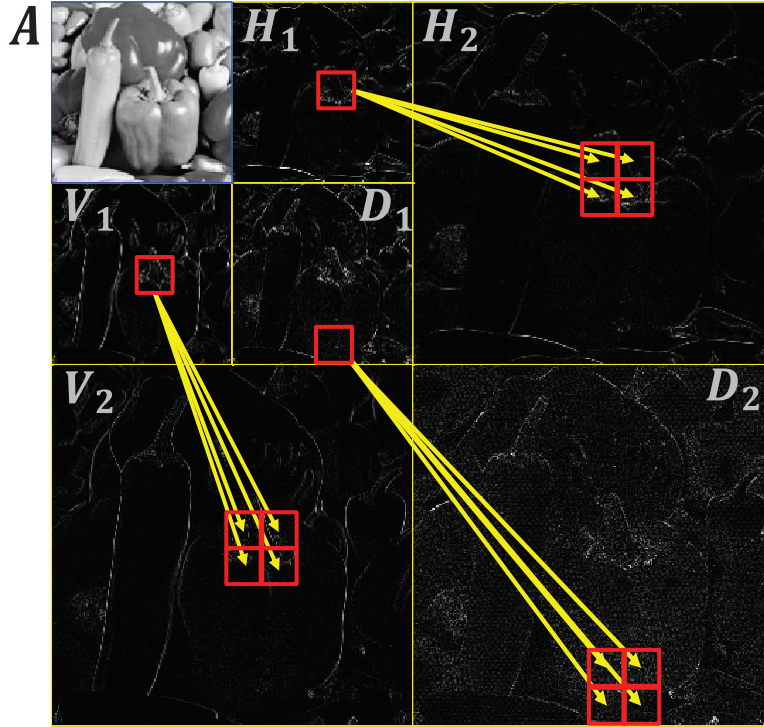


Figure 5.2: The two-dimensional wavelet transform representing an image in terms of approximate coefficients (A), and wavelet coefficients in horizontal (H), vertical (V), and diagonal (D) directions. The wavelet coefficients form quad trees with each parent coefficient having four children in the finer scale.

Let $\theta|_J = [\theta_{j,k}^b]$ represent the coefficients of the two-dimensional discrete wavelet transform up to scale J of an image represented by x . For a coefficient $\theta_{j,k}^b$, $b \in \{A, H, V, D\}$, where A, H, V, D stand for the approximate coefficients, horizontal, vertical, and diagonal subband coefficients, respectively, $j = 0, \dots, J$ represents the scale of the coefficient with $j = 0$ indicating that the coefficient is an approximate coefficient, and $k = 1, \dots, n4^{-j-1+\max(1,j)}$ represents the index of the coefficient at direction b and scale j . Generally, the first $\alpha_0 n$ wavelet coefficients in $\theta|_J$ correspond to the approximate coefficients. The remaining $\alpha_1 n, \alpha_2 n, \dots, \alpha_j n$ coefficients, where

$\sum_{j=0}^J \alpha_j = 1$, correspond to the coefficients in wavelet scales $1, 2, \dots, J$, respectively. For example, given the image in Figure 5.2, $\alpha_0 n$ is equal to the number of coefficients in A , $\alpha_1 n$ is equal to the number of coefficients in H_1, V_1 , and D_1 combined, and finally, $\alpha_2 n$ is equal to the number of coefficients in H_2, V_2 , and D_2 combined. For a two-dimensional wavelet transform with J scales, we have $\alpha_0 = 4^{-J}$ and $\alpha_j = 3\alpha_0 4^{j-1}$ for $j = 1, \dots, J$. The coefficients at scale 1 are called the *coarsest coefficients* because they only represent a rough estimate of the image.

In [36, 89] a complete analysis of the properties of wavelet coefficients is provided, and for them a hidden Markov tree (HMT) model has been developed. According to this model, every wavelet coefficient $\theta_{j,k}^b$ corresponds to a hidden state variable, which can be in either state small (\mathcal{S}) or large (\mathcal{L}). The HMT model suggests that a tree structure exists among state variables and each coefficient $\theta_{j,k}^b$ for $j = 1, \dots, J$ has a two-state mixture Gaussian probability density function (pdf) given by

$$f(\theta_{j,k}^b) = \pi_{j,k}^b \mathcal{N}(0, \sigma_{\mathcal{S},\{j,k\}}^b)^2 + (1 - \pi_{j,k}^b) \mathcal{N}(0, \sigma_{\mathcal{L},\{j,k\}}^b)^2, \quad (5.1)$$

where $\mathcal{N}(0, \sigma^2)$ represents a zero-mean Gaussian distribution with variance σ^2 , and $\pi_{j,k}^b$ is the probability that the state of $\theta_{j,k}^b$ is small (in the statistical sense) and we have $\sigma_{\mathcal{S},\{j,k\}}^b \ll \sigma_{\mathcal{L},\{j,k\}}^b$.

The state dependency between $\theta_{j,k}^b$ and its parent $\theta_{j-1, \lceil k/4 \rceil}^b$ is modeled by a state transition probability matrix $\mathbf{A}_{j,k}^b$ given by

$$\mathbf{A}_{j,k}^b = \begin{bmatrix} p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{S}} & p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{L}} \\ p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{S}} & p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{L}} \end{bmatrix}, \quad (5.2)$$

with $p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{L}} = 1 - p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{S}}$ and $p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{S}} = 1 - p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{L}}$, where $p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{S}}$ is the probability that $\theta_{j,k}^b$ is in state \mathcal{S} given its parent is in state \mathcal{S} and $p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{L}}$ is the probability that $\theta_{j,k}^b$ is in state \mathcal{L} given its parent is in state \mathcal{L} . Using $\mathbf{A}_{j,k}^b$, we can formulate $\pi_{j,k}^b$ based on the state of its parent as follows,

$$\pi_{j,k}^b = \pi_{j-1, \lceil k/4 \rceil}^b p_{j,k,b}^{\mathcal{S} \rightarrow \mathcal{S}} + (1 - \pi_{j-1, \lceil k/4 \rceil}^b) p_{j,k,b}^{\mathcal{L} \rightarrow \mathcal{S}}, \quad (5.3)$$

given the probability of being small at root coefficients ($\pi_{1,k}^b$) is known for all values of b, k . Later in this section, we will determine $\pi_{1,k}^b$ according to the image size and the number of wavelet scales.

Although the HMT model is very powerful in capturing the properties of wavelet coefficients, it requires at least $4n$ parameters to be specified. However, in [36], the authors proposed to reduce the number of parameters to $4J$ by assuming all coefficients within one wavelet scale have similar statistical parameters. Therefore, the parameters reduce to

$$\pi_{j,k}^b = \pi_j, \quad \sigma_{\mathcal{S},\{j,k\}}^b = \sigma_{\mathcal{S},j}^2, \quad \sigma_{\mathcal{L},\{j,k\}}^b = \sigma_{\mathcal{L},j}^2, \quad \text{and} \quad \mathbf{A}_{j,k}^b = \mathbf{A}_j, \quad (5.4)$$

for all b, j, k . These parameters can be estimated using a set of training images and exploiting the Expectation-Maximization (EM) algorithm as done in [36]. However, it has been shown in [89] that leveraging additional wavelet-domain image structure (such as exponential decay across scale), a reduced-parameter HMT model can be developed that is represented with only 9 meta parameters independent of the size of the image and the number of wavelet scales. Further, it has been shown in [89] that these 9 parameters take similar values for real-world images, allowing to fix a set of universal set of parameters, resulting in a *universal HMT (uHMT)*. Employing uHMT, the image-specific training is avoided¹.

Let the uHMT model be represented by a hyper-parameter Λ , which is defined as follows,

$$\Lambda = [\alpha_S, \alpha_L, C_{\sigma_S}, C_{\sigma_L}, \gamma_S, \gamma_L, C_{SS}, C_{LL}, \pi_1^l].$$

These 9 parameters are used to determine *a priori* pdfs (as given by (5.1)) for all the wavelet

¹The uHMT parameters are accurate when $J \leq \log_4 n - 3$ [89], which complies with our proposed sensing scheme.

coefficients. The variances of large and small coefficients vary over the scales [89]:

$$\sigma_{\mathcal{L},j}^2 = C_{\sigma_{\mathcal{L}}} 2^{-(J_d+j)\alpha_{\mathcal{L}}}, \quad (5.5a)$$

$$\sigma_{\mathcal{S},j}^2 = C_{\sigma_{\mathcal{S}}} 2^{-(J_d+j)\alpha_{\mathcal{S}}}, \quad (5.5b)$$

where $C_{\sigma_{\mathcal{L}}} \gg C_{\sigma_{\mathcal{S}}}$ and $J_d = \log_4 n - J$. Four parameters $C_{\sigma_{\mathcal{L}}}$, $\alpha_{\mathcal{L}}$, $C_{\sigma_{\mathcal{S}}}$, and $\alpha_{\mathcal{S}}$ characterize the variances in the marginal densities of the wavelet coefficients [89]. Using the uHMT model, \mathbf{A}_j is given by [89],

$$\mathbf{A}_j = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & j \leq 3 - J_d, \\ \begin{bmatrix} 1 - C_{\mathcal{S}\mathcal{S}} 2^{-(J_d+j)\gamma_{\mathcal{S}}} & C_{\mathcal{S}\mathcal{S}} 2^{-(J_d+j)\gamma_{\mathcal{S}}} \\ \frac{1}{2} - C_{\mathcal{L}\mathcal{L}} 2^{-(J_d+j)\gamma_{\mathcal{L}}} & \frac{1}{2} + C_{\mathcal{L}\mathcal{L}} 2^{-(J_d+j)\gamma_{\mathcal{L}}} \end{bmatrix} & j > 3 - J_d. \end{cases} \quad (5.6)$$

Unlike [89], we perform the wavelet decomposition up to an arbitrary scale J and not necessarily up to scale $\log_4 n$ (which is the full wavelet decomposition). Therefore, we add the deficit term $J_d = \log_4 n - J$ in Equations (5.5)-(5.6) to compensate for this incomplete decomposition.

The parameter π'_1 in $\mathbf{\Lambda}$ is the probability that the root coefficients are small when the wavelet decomposition is carried out up to the *coarsest* scale (i.e., $J = \log_4 n$). In this case, we find the probability that a coefficient at scale j is in the small state, π'_j , using the following recursive formula

$$\pi'_j = \pi'_{j-1} p'_j{}^{\mathcal{S} \rightarrow \mathcal{S}} + (1 - \pi'_{j-1}) p'_j{}^{\mathcal{L} \rightarrow \mathcal{S}}, \quad (5.7)$$

where $p'_j{}^{\mathcal{S} \rightarrow \mathcal{S}}$ and $p'_j{}^{\mathcal{L} \rightarrow \mathcal{S}}$ are entries of transition matrix \mathbf{A}_j given $J = \log_4 n$. In our problem, since we perform partial wavelet decomposition (i.e., $J < \log_4 n$), we set $\pi_1 = \pi'_{\log_4 n - J + 1}$ to compensate for the scale deficit. From π_1 , we can determine π_j 's using (5.3) and (5.6) for $j = 2, \dots, J$.

From (5.3) and (5.6), we note that π_j is an increasing function of j . Therefore, the sparsity rate of wavelet coefficients at scale j , given by $s_j = 1 - \pi_j$, is a decreasing function of j . In the next section, we will exploit this *non-uniform sparsity* property of wavelet coefficients to design a novel CS measurement matrix.

As an example, for a 128×128 image with $J = 4$ and using the parameters of uHMT model from [89] ($\gamma_S = \gamma_L = 1$, $C_{SS} = 0.2$, $C_{LL} = 0.4$, $\alpha_S = \alpha_L = 2.5$ and $\pi'_1 = 0.5$), we find $\pi_1 = 0.731$, $\pi_2 = 0.858$, $\pi_3 = 0.925$ and $\pi_4 = 0.961$. Therefore, the sparsity rates of different scales are found as $s_1 = 0.269$, $s_2 = 0.142$, $s_3 = 0.075$, $s_4 = 0.039$. The overall sparsity rate of the wavelet coefficients excluding the approximate coefficients is found as $\sum_{j=1}^J \alpha_j s_j = 0.055$ and we see the sparsity rate of lower scales are much higher than the overall sparsity rate of the image.

Model-based Non-Uniform Compressed Sampling

In this section, we introduce our proposed *non-uniform* measurement matrix Φ_{uHMT} for natural images. Our approach is based on the integration of the uHMT model of natural images' wavelet coefficients into the design of the measurement matrix.

Non-Uniform Sampling

It is known that the wavelet coefficients of an image show an *exponential decay* along the scales of the wavelet tree [36, 89]. This means that most of images' energy is carried by the approximate coefficients and the wavelet coefficients of the coarser scales. In addition, the sparsity rate of wavelet coefficients at scale j , given by $s_j = 1 - \pi_j$, is a decreasing function of j . Therefore, it seems rational to sample the coefficients at coarser scales, which are the initial coefficients in $\theta|_J$, with a higher probability and decrease the probability that a coefficient is sampled by a CS measurement as we increase j . For these purposes, we propose a novel measurement matrix Φ_{uHMT} that integrates the uHMT model of wavelet coefficients into its design.

First, we *directly* sample the $\alpha_0 n = 4^{-J} n$ approximate coefficients. Next, we nonuniformly sample the remaining wavelet coefficients using a *nonuniformly sparse matrix* $\Phi_{NU} = [\Phi_1, \Phi_2, \dots, \Phi_J]$,

where sub-matrix Φ_j corresponds to the $\alpha_j n = 3 \times 4^{-(J-j+1)} n$ wavelet coefficients at scale j , for $j = 1, 2, \dots, J$. Here $J \leq \log_4 n$ represents the depth of the wavelet trees (i.e., the level up to which the wavelet decomposition has been done). In addition, we set every row of Φ_{NU} to have L non-zero coefficient, chosen from a zero mean, unit variance Gaussian distribution. The structure of the proposed measurement matrix Φ_{uHMT} is shown in Figure 5.3, in which $I_{\alpha_0 n}$ is an identity matrix of size $\alpha_0 n \times \alpha_0 n$.

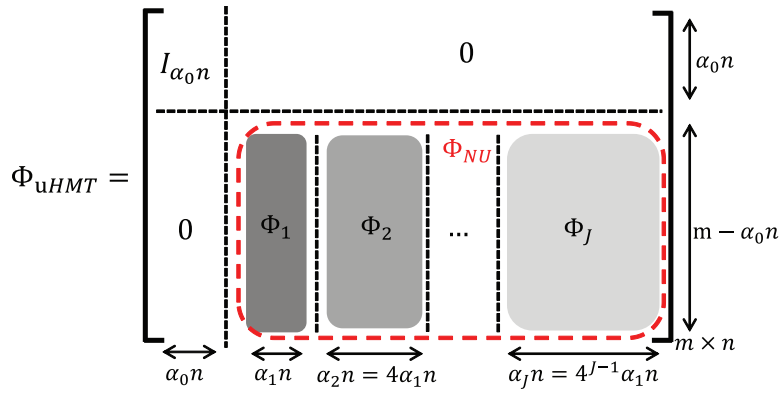


Figure 5.3: The proposed non-uniform measurement matrix for compressive sensing of wavelet coefficients of an image. A darker color corresponds to a denser matrix.

If we take m CS measurements, $m - \alpha_0 n$ of them will be generated nonuniformly from the wavelet coefficients other than the approximate coefficients. To impose the non-uniform sampling through Φ_{NU} , we set each row of Φ_j to have L_j non-zero entries and $\sum_{j=1}^J L_j = L$. The value of L is proportional to n , where $L = \frac{n}{c}$ and c is a large constant.

Let $\mathbf{y} = \Phi_{uHMT} \boldsymbol{\theta}|_J$ denote the compressive samples of wavelet coefficients of an image. For a measurement y_i ($\alpha_0 n + 1 \leq i \leq m$), the number of contributing coefficients from scale j is L_j .

Since the significance of wavelet coefficients decreases as j increases, we set

$$\frac{L_1}{\alpha_1} > \frac{L_2}{\alpha_2} > \dots > \frac{L_J}{\alpha_J}. \quad (5.8)$$

This will ensure that the coefficients at lower scales contribute to more measurements and correspondingly they have more contribution in each measurement.

In our proposed scheme, we relate L_j to the sparsity of wavelet coefficients as follows:

$$L_j = \begin{cases} \left\lceil \frac{\alpha_j s_j}{\sum_{i=1}^{J-1} \alpha_i s_i} L \right\rceil & j=1, \dots, J-1 \\ L - \sum_{i=1}^{J-1} L_i & j=J. \end{cases}, \quad (5.9)$$

where $s_j = 1 - \pi_j$ is the sparsity rate of the wavelet coefficients at scale j and π_j is found using (5.3). In (5.9), L_j will be greater than zero if we have $L \geq \frac{\sum_{j=1}^J \alpha_j s_j}{\alpha_J s_J} (J - 1)$. Since $J = \mathcal{O}(\log n)$ represents the levels of the wavelet transform, choosing L to be at least $\mathcal{O}(\log n)$ suffices to have $L_j \geq 0$. As we mentioned in the proposed non-uniform sampling, we set $L = \frac{n}{c}$. Therefore for sufficiently large n , L_j will be greater than or equal to zero.

Mutual Coherence of Φ_{uHMT} and Ψ

In order to have an efficient sampling and recovery process, the measurement matrix and the sparsifying basis must be mutually incoherent. The *mutual coherence* μ of the measurement matrix and the sparsifying matrix is defined as $\mu(\Phi, \Psi) = \sqrt{n} \max_{k,j} |\langle \phi_k, \psi_j \rangle|$ [96], where ϕ_k is the k^{th} row Φ and ψ_j is the j^{th} column of Ψ and $\|\phi_k\|_2 = 1$ for all $k = 1, \dots, m$ and $\|\psi_j\|_2 = 1$ for all $j = 1, 2, \dots, n$.

As shown in Figure 5.3, our proposed measurement matrix can be written as

$$\Phi_{uHMT} = \begin{bmatrix} [\mathbf{I}_{\alpha_0 n} \mid \mathbf{0}]_{\alpha_0 n \times n} \\ [\mathbf{0} \mid \Phi_{NU}]_{(m-\alpha_0 n) \times n} \end{bmatrix}.$$

Therefore, the mutual coherence of Φ_{uHMT} and Ψ is

$$\mu(\Phi_{uHMT}, \Psi) = \max\{\mu([\mathbf{I}_{\alpha_0 n} \mid \mathbf{0}], \Psi), \mu([\mathbf{0} \mid \Phi_{NU}], \Psi)\}. \quad (5.10)$$

Given Ψ is an orthonormal basis with $|\psi_{ij}| = \mathcal{O}(\frac{\log n}{\sqrt{n}})$ for all $i, j = 1, \dots, n$, we have

$$\mu([\mathbf{I}_{\alpha_0 n} \mid \mathbf{0}], \Psi) = \mathcal{O}(\log n).$$

In addition using the following lemma, we find the mutual coherence of $[\mathbf{0} \mid \Phi_{NU}]$ and any orthonormal bases.

Lemma 3. *Let Φ be an m -by- n sparse random matrix. Assume every row of Φ has $L = \frac{n}{c}$ iid non-zero entries (not necessarily uniformly distributed) chosen from $\mathcal{N}(0, \frac{1}{L})$ and c is a large constant. For any arbitrary orthonormal basis $\Psi_{n \times n}$, we have $\Xi = \Phi\Psi$ is an iid zero-mean Gaussian matrix with variance less than or equal to $\frac{1}{L}$.*

Proof. Assume ϕ_k is the k^{th} row Φ and ψ_j is the j^{th} column of Ψ and the L indices of non-zero elements in ϕ_k are denoted by set R_k ($R_k \subset \{1, \dots, n\}$). The matrix $\Xi = \Phi\Psi$ with entries $\Xi_{kj} = \sum_{i=1}^n \phi_{ki}\psi_{ij} = \sum_{i \in R_k} \phi_{ki}\psi_{ij}$ is defined. Therefore, it can be easily seen that for all $j = 1, \dots, n$ and $k = 1, \dots, m$, Ξ_{kj} is a linear combination of L Gaussian random variables. Accordingly, Ξ_{kj} itself is a Gaussian random variable with mean $E\{\Xi_{kj}\} = \sum_{i \in R_k} E\{\phi_{ki}\}\psi_{ij}$ and variance $\text{Var}(\Xi_{kj}) = \sum_{i \in R_k} \psi_{ij}^2 \text{Var}(\phi_{ki})$. Hence, we have the following,

$$E\{\Xi_{kj}\} = 0,$$

$$\text{Var}(\Xi_{kj}) = \frac{1}{L} \sum_{i \in R_k} \psi_{ij}^2 \leq \frac{1}{L}$$

□

Lemma 3 implies that $[\mathbf{0}|\Phi_{NU}] \times \Psi$ is a Gaussian random matrix with iid zero-mean entries and variance less than or equal to $1/L$. Using the union bound for the maximum absolute magnitude of a Gaussian matrix, $|\Xi_{kj}|$ for all $j = 1, \dots, n$ and $k = 1, \dots, m$ can be bounded as follows [93],

$$P \left(\max_{1 \leq k \leq m, 1 \leq j \leq n} |\Xi_{kj}| \geq t \right) \preceq 2nm \exp \left(-\frac{t^2}{2\sigma^2} \right) \leq 2n^2 \exp \left(-\frac{t^2}{2\sigma^2} \right), \quad (5.11)$$

where $\sigma^2 \leq \frac{1}{L} = \frac{c}{n}$ and \preceq represents *asymptotically smaller than or equal*. Choosing $t = \sqrt{\frac{2c \log \left(\frac{2n^2}{\delta} \right)}{n}}$, the inequality in (5.11) becomes,

$$P \left(\max_{1 \leq k \leq m, 1 \leq j \leq n} |\Xi_{kj}| \leq \sqrt{\frac{2c \log \left(\frac{2n^2}{\delta} \right)}{n}} \right) \succeq 1 - \delta. \quad (5.12)$$

Inequality (5.12) shows that $\mu([\mathbf{0}|\Phi_{NU}], \Psi) = \mathcal{O} \left(\sqrt{\log \left(\frac{n}{\sqrt{\delta}} \right)} \right)$, with probability at least $1 - \delta$. Consequently, $\mu(\Phi_{uHMT}, \Psi) = \max \{ \mu([\mathbf{I}_{\alpha_0 n}|\mathbf{0}], \Psi), \mu([\mathbf{0}|\Phi_{NU}], \Psi) \} = \mathcal{O} \left(\sqrt{\log \left(\frac{n}{\sqrt{\delta}} \right)} \right)$, which is close to optimal bound except for the $\log n$ factor. It is worth mentioning that the Haar wavelet satisfies the condition $|\psi_{ij}| = \mathcal{O} \left(\frac{\log n}{\sqrt{n}} \right)$ for all $i, j = 1, \dots, n$ [97]. Therefore, the Haar wavelet basis, Ψ_{Haar} , satisfies $\mu(\Phi_{uHMT}, \Psi_{Haar}) = \mathcal{O} \left(\sqrt{\log \left(\frac{n}{\sqrt{\delta}} \right)} \right)$.

Algorithm 1 summarizes our proposed algorithm for non-uniform compressive sampling of 2D wavelet coefficients of an image.

Algorithm 1 Non-Uniform compressive sampling of wavelet coefficients of an image.

- 1: Initialize $\alpha_0 = 4^{-J}$, $\alpha_j = 3 \times 4^{-(J-j+1)}$, $s_j = 1 - \pi_j$ for $j = 1, \dots, J$. The values of π_j 's are found using (5.3).
 - 2: Generate $\Phi_{NU} = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_J]$ based on the structure in Figure 5.3. Each row of Φ_j has L_j non-zero entries, which are chosen as iid zero-mean and unit variance Gaussian random variables.
 - 3: Directly sample $\alpha_0 n$ initial coefficients of θ .
 - 4: Sample the remaining $(1 - \alpha_0)n$ coefficients of θ using the updated Φ_{NU} .
-

Model-Based Non-Uniform CS Recovery

In previous section, we integrated the uHMT signal model into the design of a novel non-uniform measurement matrix. In this section, we utilize the same model at the recovery step. We consider two Bayesian recovery algorithms (i.e., the CSBP algorithm [7] and the AMP algorithm [43]) and modify them to exploit the uHMT model.

CSBP-uHMT: Integrating the uHMT Model into CSBP

One of the advantages of the CSBP recovery algorithm [7] is its ability to accommodate *a priori* knowledge about the signal model in the CS recovery process. In the conventional CSBP, all the variable nodes are assigned *the same a priori* pdf that considers the *sparsity rate* $s = k/n$ as the probability that each coefficient is at the large state.

In contrast, in our proposed CSBP-uHMT, we assign a *different* prior to each variable node based on the uHMT model. Specifically, each wavelet coefficient $\theta_{j,k}^b$ receives an *a priori* mixture Gaussian pdf

$$f(\theta_{j,k}^b) = \pi_j \mathcal{N}(0, \sigma_{S,j}^2) + (1 - \pi_j) \mathcal{N}(0, \sigma_{L,j}^2). \quad (5.13)$$

The hyper-parameter $\Lambda = [\alpha_S, \alpha_L, C_{\sigma_S}, C_{\sigma_L}, \gamma_S, \gamma_L, C_{SS}, C_{LL}, \pi'_1]$ is used to determine π_j , $\sigma_{S,j}^2$, and $\sigma_{L,j}^2$ using (5.1)-(5.7).

AMP-uHMT: Integrating the uHMT into AMP

The AMP algorithm [43] and its model-based version (Turbo AMP) [2] are proven to have very competitive recovery performances and very low computational complexity. Unlike Turbo AMP that employs a learning-based approach to obtain the statistical parameters of HMT model, we directly apply the uHMT parameters to the AMP algorithm. In Turbo AMP, the variances $\sigma_{\mathcal{S},\{j,k\}}^b$ ² and $\sigma_{\mathcal{L},\{j,k\}}^b$ ² and the probabilities $\pi_{j,k}^b$ are assumed to be random variables with known distributions [2, Eq. (3)-(8)]. In our proposed work we treat them as fixed and known parameters found by uHMT model using Equations (5.3) and (5.5).

Clearly the uHMT model may not be as accurate as the training-based approaches. However, it has been shown in [89] that using uHMT has a negligible degrading effect on the accuracy of images.

Simulation Results and Discussion

In this section, we compare the recovery performance of our proposed model-based uHMT-NCS algorithms (which is based on the integration of the HMT model into both sampling and recovery steps) with other state-of-the-art model-based CS recovery schemes that only integrate the model at the recovery step and use conventional CS at the sampling step. The algorithms from the literature that we have chosen are Turbo AMP [2], model-based CS [3], and TSWCS-MCMC [4]. To implement these algorithms, we use the full Gaussian measurement matrices with iid entries with these algorithms (See Table 5.1). We consider a 128×128 image ($n = 16,384$) for our simulations (given in Figures 5.7(a) and 5.8(a)). The sparsifying basis, Ψ , is considered to be the two-dimensional Haar wavelet basis. The parameters of uHMT model are set as described in [89] (we verified these parameters for a set of 128×128 test natural images and Haar wavelet): $\alpha_{\mathcal{S}} = \alpha_{\mathcal{L}} = 2.5$, $C_{\sigma_{\mathcal{S}}} = 2^7$, $C_{\sigma_{\mathcal{L}}} = 2^{13}$, $\gamma_{\mathcal{S}} = \gamma_{\mathcal{L}} = 1$, $C_{\mathcal{S}\mathcal{S}} = 0.2$, $C_{\mathcal{L}\mathcal{L}} = 0.4$, and $\pi'_1 = 0.5$. We set

$L = 40$ and $J = 3$. Therefore, we find $\pi_1 = 0.858$, $\pi_2 = 0.925$, and $\pi_3 = 0.961$. For uHMT-NCS simulations, the matrix Φ is generated using Algorithm 1.

In Figure 5.4, we have shown the normalized recovery error (NRE = $\frac{\|\hat{\theta} - \theta\|_2}{\|\theta\|_2}$) versus the number of measurements. We depicted the performance of our proposed uHMT-NCS (with CSBP-uHMT and AMP-uHMT) and Turbo AMP, model-based CS, and TSWCS-MCMC schemes. For these algorithms, the tunable parameters such as wavelet levels are set for the best performance. As we see in Figure 5.4, Bayesian recovery-based algorithms such as TSWCS-MCMC [4] and our proposed uHMT-NCS have very good performances even with small number of measurements. Among all schemes uHMT-NCS with AMP-uHMT recovery performs the best.

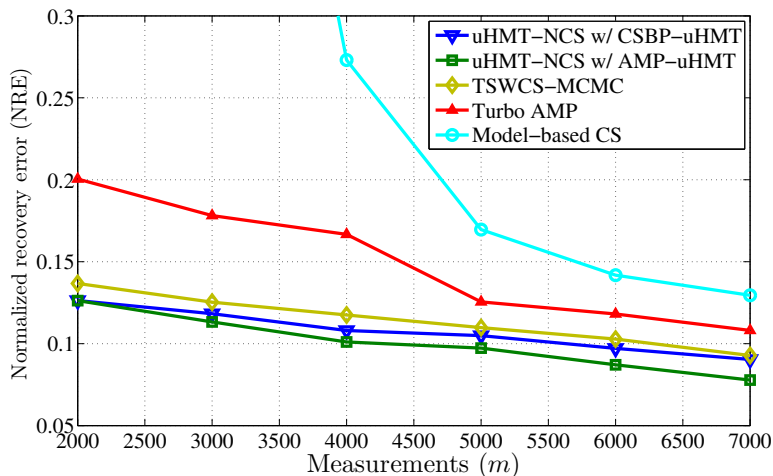


Figure 5.4: Comparison between the recovery performances of our proposed schemes (uHMT-NCS w/ CSBP-uHMT and uHMT-NCS w/ AMP-uHMT) and different CS recovery algorithms (Turbo AMP [2], Model-based CS [3], and TSWCS-MCMC [4]).

A similar simulation is performed for 27 sample images (see Figure 5.5) from the Microsoft object

class recognition database v2². We applied our proposed uHMT-NCS (with CSBP-uHMT and AMP-uHMT) as well as Turbo AMP, model-based CS, and TSWCS-MCMC to these images. The NRE performance of each image using each algorithm is shown in Figure 5.6. For this simulation we set $m = 4000$ and $J = 3$ and 2D Haar wavelet in our own algorithms. However, the model-based CS algorithm is tuned for its best performance which is achieved with $J = 6$.



Figure 5.5: A set of 27 sample images from Microsoft object class recognition database. The numbers in each image are solely for referencing and are not part of the image.

²All images are cropped to be rectangular, and resized to 128×128 . The database is available for download at ["http://research.microsoft.com/en-us/projects/objectclassrecognition/"](http://research.microsoft.com/en-us/projects/objectclassrecognition/)

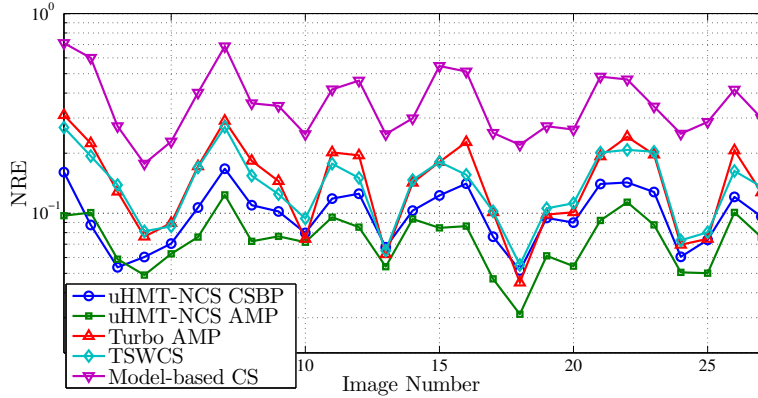


Figure 5.6: NRE performance for the sample images in Figure 5.5.

As we see in Figure 5.6, in 25 out of 27 images our proposed uHMT-NCS performs better than other algorithms in comparison.

Figures 5.7 and 5.8 provide the visual comparison of the recovery performance of uHMT-NCS with state-of-the-art algorithms. The number of measurements in Figure 5.7 and Figure 5.8 are $m = 4000$ and $m = 6000$, respectively. Figures 5.7(a) and 5.8(a) show the original image, Figures 5.7(b)-5.8(b), 5.7(c)-5.8(c), Figures 5.7(d)-5.8(d), and Figures 5.7(e)-5.8(e) show the visual performance of model-based CS [3], visually weighted CS [92], TSWCS-MCMC Bayesian algorithm [4], and Turbo AMP [2], respectively. Finally, Figures 5.7(f)-5.8(f), and 5.7(g)-5.8(g) represent our proposed uHMT-NCS algorithm with CSBP-uHMT and AMP-uHMT recovery algorithms, respectively. As we see, uHMT-NCS with both CSBP-uHMT and AMP-uHMT recovery algorithms has a significantly smaller visual distortion compared to the others.

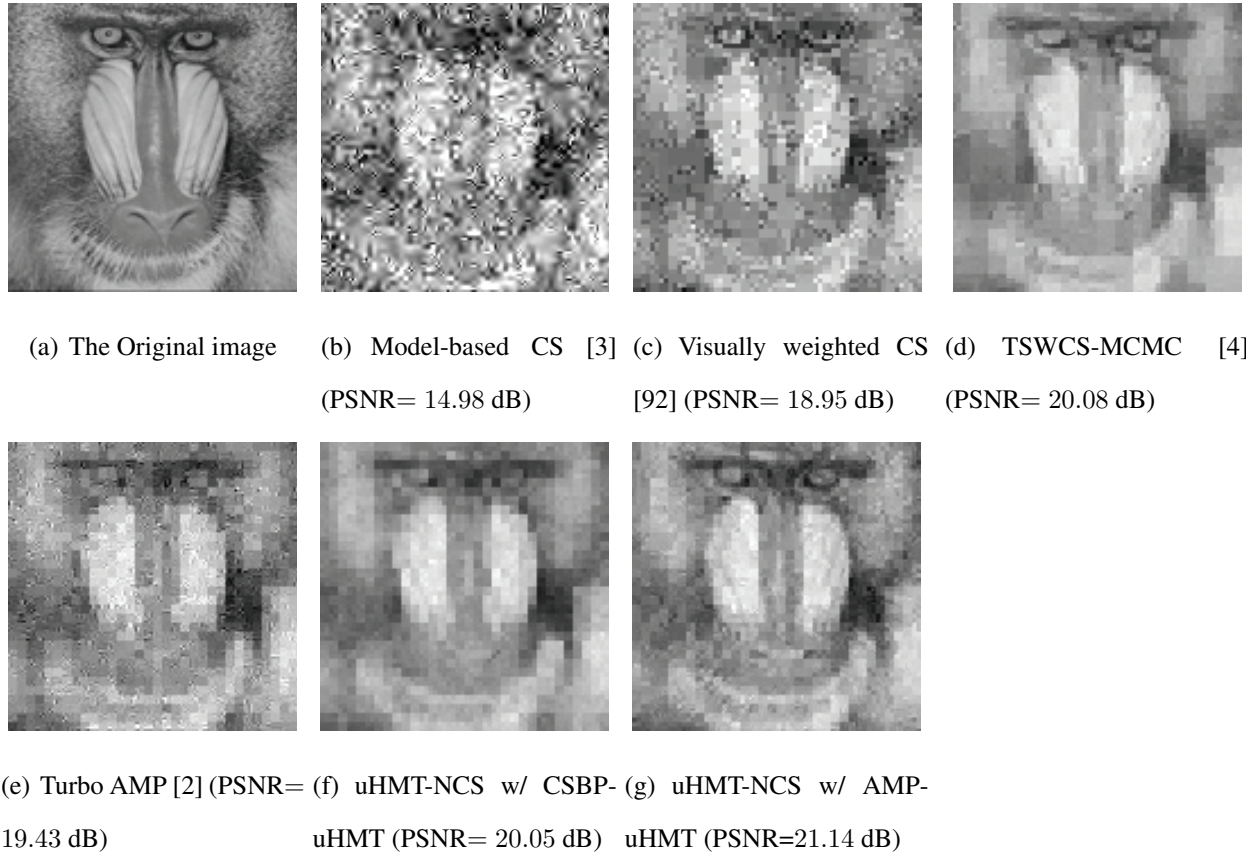


Figure 5.7: Comparing the visual performance of different CS schemes at $m = 4000$ measurements.

In the next simulation, we investigate the effect of employing our proposed non-uniform Φ_{uHMT} in the overall performance of the proposed uHMT-NCS algorithm. In other words, we would like to see how much of the improved performance is due to the integration of the model at the sampling step through our proposed non-uniform Φ_{uHMT} . In Figure 5.9, we compare the NRE performance of AMP [43] with the cases when the uHMT model is added to only sampling step (uHMT-NCS w/ AMP), only recovery step (sparse random sampling w/ AMP-uHMT), and both steps (uHMT-NCS w/ AMP-uHMT) (See Table 5.1). When uHMT is only applied to the recovery step the measurement matrix is sparse random matrix with $L = 40$ non-zero entries per row.

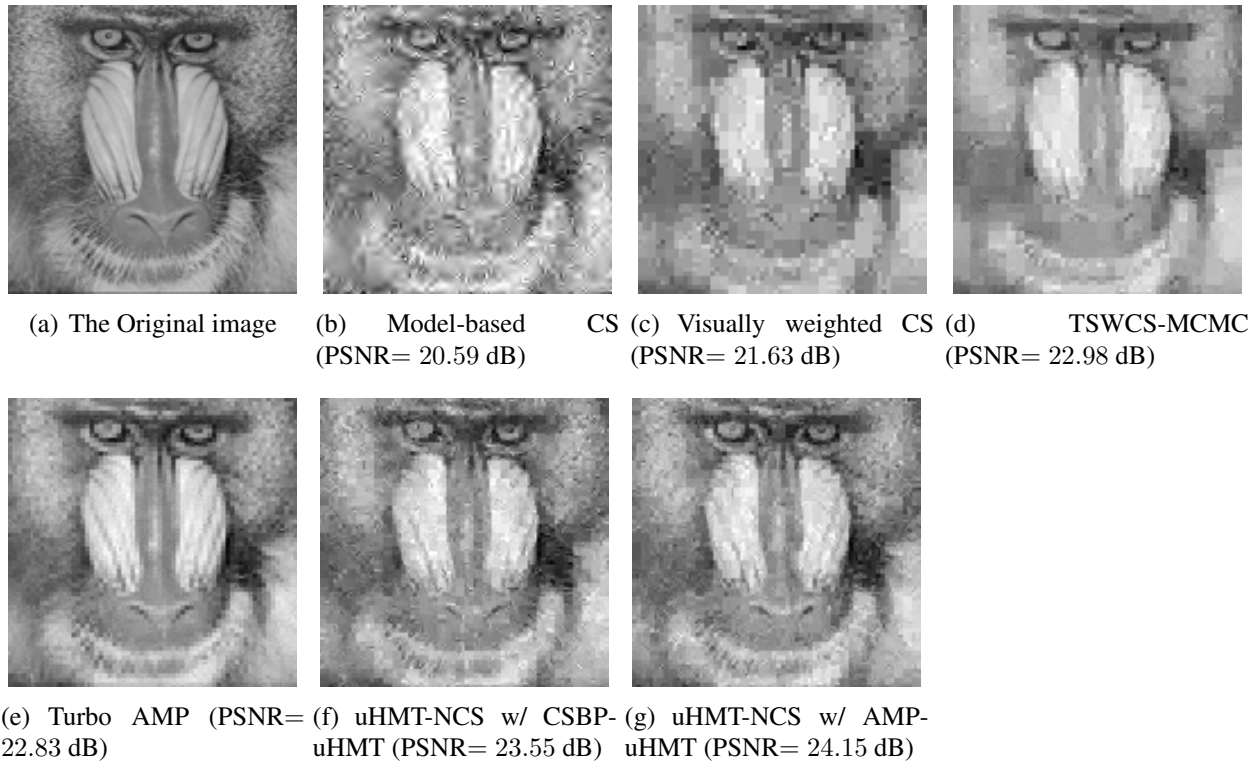


Figure 5.8: Comparing the visual performance of different CS schemes at $m = 6000$ measurements.

As we see in Figure 5.9, including the uHMT model only at the sampling step (uHMT-NCS w/ AMP) improves the performance compared to the case when it is solely added to the recovery step. Therefore, *integrating the model at the sampling phase* is even *more effective* than such integration at the *recovery phase*. Clearly, when uHMT model is included in both sampling and recovery steps (denoted by uHMT-NCS w/ AMP-uHMT) the most performance improvements is achieved.

Table 5.1: Properties of different CS sampling and recovery schemes

Scheme	Model-based Recovery	Model-Based Sampling	Measurement Matrix
Model-based CS [3]	✓	x	Full Gaussian
TSWCS [4]	✓	x	Direct sampling+ full Gaussian
Turbo AMP [2]	✓	x	Full Gaussian
AMP [43] w/ sparse random Φ	x	x	Sparse random Φ ($L = 40$)
uHMT-NCS w/ AMP	x	✓	Φ_{uHMT}
Sparse Random Sampling w/ AMP-uHMT	✓	x	Sparse random Φ ($L = 40$)
uHMT-NCS w/ AMP-uHMT	✓	✓	Φ_{uHMT}

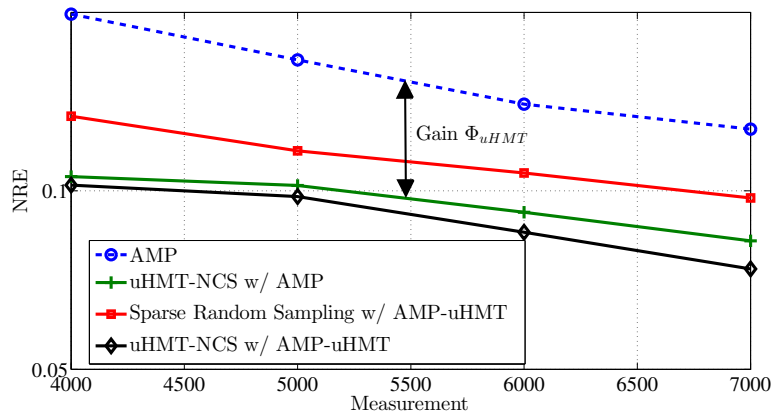


Figure 5.9: Improvement over the conventional AMP technique by exploiting the uHMT signal model at the sampling step (uHMT-NCS w/ AMP), the recovery step (sparse random sampling w/ AMP-uHMT), and both (uHMT-NCS w/ AMP-uHMT).

In the next simulation, we consider the effect of changing the row weights of Φ_j s, on the performance of the uHMT-NCS. In this simulation, we set $m = 4000$, $L = 40$ and $J = 3$. Accordingly, we find the NRE of uHMT-NCS with CSBP-uHMT changing L_1 and L_2 parameters. As shown in Figure 5.10, employing the values of L_1 and L_2 given in (5.9) (i.e. $L_1 = 6$ and $L_2 = 12$) results in a smaller NRE which is about 6% better than the uniform sampling case (i.e. $L_1 = 2$, $L_2 = 8$). However, for this particular example for $4 \leq L_1 \leq 6$ and $10 \leq L_2 \leq 18$ the results are within the 2% of its minimum value.

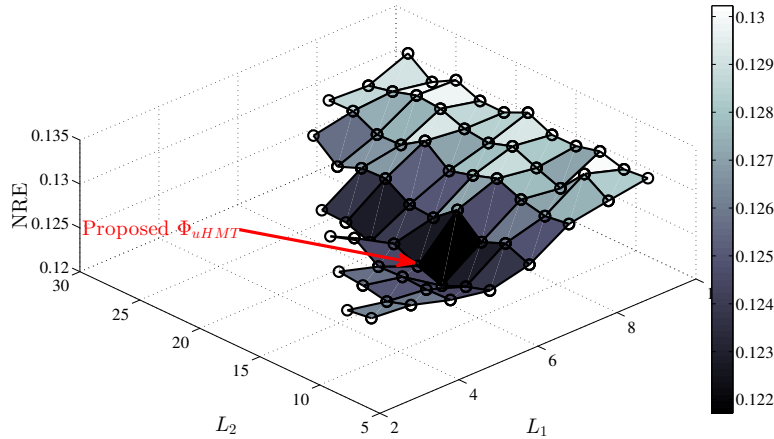


Figure 5.10: NRE versus L_1 and L_2 . Choosing L_1 and L_2 using (5.9) results in almost 6% performance improvement.

Finally, we study the effect of the parameter J (the depth of wavelet trees) in the performance. In Figure 5.11, we have depicted the NRE of our proposed uHMT-NCS algorithm with the CSBP-uHMT recovery scheme versus J for the sample image of Figure 5.7(a) for different values of m .

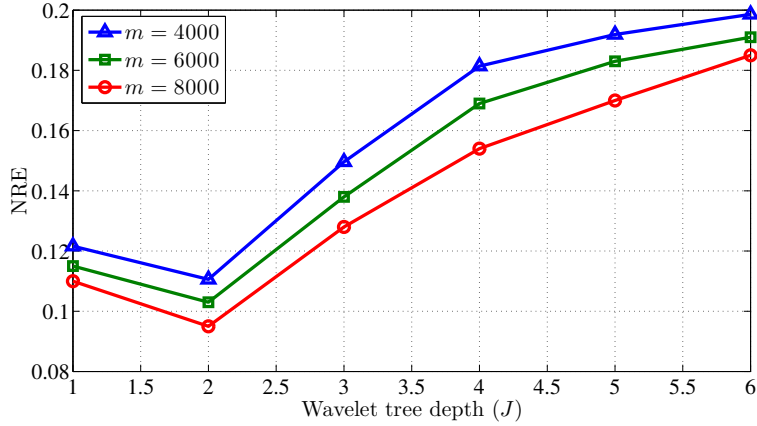


Figure 5.11: Recovery performance of uHMT-NCS with CSBP-uHMT recovery versus different depths of the wavelet tree J .

As shown, by increasing tree depth, NRE initially decreases and then increases. The optimal performance is obtained at $J = 2$. This observation can be explained using the *exponential decay* property of wavelet coefficients. This property suggests that most of images energy are located in the initial signal coefficients. When $J = 1$, direct sampling dominates the measurements. Therefore, many coefficients will not be included in any measurement and that results in poor performance. As J increases, fewer coefficients are measured directly and more non-uniform CS samples are generated. However, by increasing J the sparsity rate of higher scales decreases exponentially and it is inefficient to allocate too many measurements to sample these coefficients. We used optimal value $J = 2$ for our previous simulations.

Conclusion

In this chapter, we have developed a *model-based CS non-uniform sampling and recovery* scheme (uHMT-NCS) for natural images that exploits the universal hidden Markov tree (uHMT) model of

wavelet coefficients in *both* CS sampling and CS recovery steps. As we have shown, not only the signal model can be utilized to optimize the initial priors for Bayesian CS recovery algorithms, but it can also be employed in the design of new CS measurement matrices. The results of our numerical experiments suggest a significant performance gain compared to the state-of-the-art model-based CS algorithms. To the best of our knowledge, our work is one of the first Bayesian-based algorithms to consider the signal model in the design of the CS measurement matrix as well as the recovery scheme. Although we considered the natural images as our underlying signal, our approach is not limited to the image processing applications and the wavelet sparsifying bases. After the model extraction, similar procedure can be adopted for compressive sampling and recovery of other real-world signals.

CHAPTER 6: CLUSTER-CMSS: A CLUSTER-BASED COORDINATED SPECTRUM SENSING

A coordinated multiband spectrum sensing policy for mobile and geographically dispersed cognitive radio networks (CRNs), referred to as Cluster-CMSS, is proposed. The goal is to detect the spectrum holes and to assign each secondary user (SU) a sensing channel with the maximum probability of being empty. In geographically dispersed CRNs, channels availability varies over the space and this makes the sensing outcomes and sensing assignments location-dependent. However, if SUs are not equipped with location-finding technologies, fusing the sensing outcomes to find the optimal spectrum sensing assignments for the next sensing time becomes challenging for the base station. To tackle this problem, we introduce a metric solely based on the sensing outcomes of SUs. Using this metric along with a low-complexity clustering algorithm enables the base station to efficiently divide the network into clusters. Further, we present an adaptive learning algorithm to learn the dynamic behavior of channels occupancy in the primary network. The proposed learning algorithm considers SUs mobility model to determine the optimal learning window. To determine the sensing assignments, the base station performs a graph-theory-based coordinated multiband spectrum sensing within each cluster. Specifically, a weighted bipartite matching is employed. We have shown that Cluster-CMSS significantly increases the spectrum opportunity discovery ratio for SUs at the cost of a slight increase in the energy consumption associated with spectrum sensing.

Introduction

Cognitive radio (CR) is a promising solution to alleviate today's spectrum deficiency caused by an increased demand for the wireless technologies [8]. The CR paradigm allows a new type of users called unlicensed users or secondary users (SUs) to coexist with the licensed users or primary users

(PUs). The SUs are allowed to access the spectrum provided that they do not interfere with the PUs. The under-utilized spectrum bands that can be used by the SUs are called spectrum holes [9].

The availability of spectrum holes varies in both *time* and *space* since the PUs' activity is dispersed in both temporal and spatial domains. An ideal CR is able to efficiently detect and utilize all spectrum holes. Due to the dynamic behavior of PUs, SUs should constantly be aware of the occupancy status of multiple narrow bands or channels of spectrum (a.k.a., wideband spectrum sensing). However, implementing wideband spectrum sensing requires considerable amount of time [10] or complex hardware [11] to obtain a fairly good estimate of the entire spectrum. This lengthy estimation will significantly reduce SUs opportunity to transmit their own data [12].

The problem that we are trying to tackle in this chapter is to develop an spectrum sensing assignment policy that maximizes discovery ratio of spectrum holes while the overhead spectrum sensing time is minimized by sensing only one channel per SU at a time. Our proposed method is referred to as *cluster-based coordinated multiband spectrum sensing* (Cluster-CMSS). In this problem SUs are mobile and can communicate with a central node or *base station* (BS). This is a very complex problem with numerous challenges. The main challenges are limited ability of SUs in sensing the spectrum, geographically dispersed SU distribution, dynamic PU activity and inaccurate sensing. To the best of our knowledge, this is the first attempt that addresses all these challenges simultaneously.

Contribution of This Chapter

In this work we propose a spectrum sensing policy for geographically dispersed networks that does not require location information of SUs. Accordingly, the BS identifies the SUs with *highly correlated spectrum sensing results* and determines the SUs that most likely experience a similar set of spatial spectrum holes. To this aim, we propose a spectrum sensing policy in which the

BS groups the SUs into several clusters based on the *correlation* of their sensing outcomes and performs *a coordinated spectrum sensing* within each cluster independently. The BS then assigns each SU a channel to sense such that the assigned channel is expected to be empty with a high probability. After sensing is done, every SU that has sensed an empty PU channel will have at least one unique channel to access.

The main contribution of this chapter is addressing coordinated spectrum sensing problem in the geographically dispersed and mobile cognitive radio networks. The novelty of the proposed framework is three-fold. First, we propose a novel *metric* that allows us to group the SUs based on the similarity of spectrum holes that they can find. Second, we propose a learning algorithm for estimating the PU's dynamic based on the mobility of SUs. Third, we propose a novel energy-efficient and fast coordinated spectrum sensing policy that maximizes the channel discovery ratio for SUs.

Related Work

Spectrum sensing in cognitive radio networks is a very well-studied topic in the literature [10, 98–100]. However, some of its aspects received more attentions compared to others. For example many studies have extensively covered issues such as *cooperation* among SU to reliably detect the spectrum holes [98, 99] or spectrum sensing employing cyclostationary features [10]. On the other hand, the networks that include mobility or geographically dispersed SUs are under-investigated.

The idea of SUs' coordination to perform multiband spectrum sensing was introduced in [101]. Later, a joint coordinated spectrum sensing and access scheme for wideband CRNs was introduced in [102]. The distributed scheme in [102] utilizes a common control channel among all SUs to perform coordination, and spectrum sensing is carried out using a negotiation-based approach. The problem of spectrum opportunity discovery when the BS is aware of SUs locations is studied

in [98]. The authors quantified the gain that is achieved by simultaneously employing both spatial and temporal spectrum holes versus the employing them individually. In [99], the problem of joint spatial and temporal spectrum opportunity discovery for a case of single PU band is considered.

In [103, 104], the joint problem of spectrum sensing and access in geographically dispersed cognitive radio networks are formulated in form of a Restless Multi Armed Bandit (RMAB) problem and the bounds for the regret of the proposed policies have been found. In [105] an iterative Hungarian algorithm is proposed to find the sensing assignment that minimizes the probability of miss-detection. This algorithm assigns SUs to sense different channels assuming that the channel availability are consistent among all SUs.

In [106], a machine-learning-aided spectrum sensing policy is proposed. In the aforementioned policy, each SU is assigned to sense the channel that provides that SU with the highest throughput.

In all the mentioned studies, it is assumed that SUs are static. Given that mobility significantly affects the performance of spectrum sensing [107], it is of great importance to consider the effect of mobility in learning the PU's activity and also spectrum sensing assignment. However, in the context of spectrum sensing for cognitive radio this problem has not received much attention. Most of the previous work on mobile cognitive radio networks are dealing with routing or connectivity issues [108]. The work in [107] is among the first ones that shows SU's mobility increases spatio-temporal diversity in the received PU's signal and improves the sensing performance. In [109], a mobility-aware cluster-based cooperative spectrum sensing approach has been proposed. The authors have shown that in case of cooperation the mobility-aware clustering improves the channel discovery ratio and the throughput.

In [110], we proposed a cluster-based coordinated spectrum sensing algorithm that employs the Kullback-Liebler divergence between the previous sensing results of SU to form the clusters. After the clusters are formed, SUs within each cluster perform the CMSS algorithm. In [111], a non-

centralized clustering approach is employed to cluster the SUs based on their channel sensing outcomes. The work in [111] assumes SUs already have the availability information of all channels either through sensing or a databases query and forms the clusters such that the cluster members have maximum idle PU channels in common. In [112–114], the SUs are grouped into clusters and for each cluster the best channel to sense is determined. In these papers, the members of a cluster all sense the same channel, which is in contrast to our approach. Additionally in [111, 115], clustering is employed to reduce the network management traffic.

Although different aspects of spectrum sensing in cognitive radios have been studied individually, to the best of our knowledge, this is the first study that addresses the cluster-based coordinated multiband spectrum sensing for mobile SUs.

System Model

The entire spectrum of interest is divided into M *orthogonal frequency subbands* or PU channels each with bandwidth W . The SU network consists of N mobile wireless terminals (or simply SUs) and a stationary BS. Each SU is equipped with a *single antenna* and can perform either sensing or transmission at a time. The RF frontend of SUs employs *energy detectors* and can reliably sense only one PU channel per sensing. For now, we consider the ideal sensing case (in which probabilities of miss-detection and false-alarm are both zero) while describing our proposed policy. Later, we consider the non-ideal sensing scenarios. In addition, similar to many other studies (e.g., [113, 116, 117]), we assume a dedicated common control channel exists between the SUs and the BS and all SUs can directly communicate with the BS. The SUs move based on a *random waypoint* mobility model. According to this model, SUs movement occurs in *epochs*. At the beginning of each epoch, an SU independently chooses a destination in the network (a *waypoint*) uniformly at random and starts moving toward that destination at a constant velocity, which is

chosen uniformly at random from the interval $[v_{min}, v_{max}]$. When SU reaches its destination, it pauses for t_p seconds until it starts a new epoch following the same rule. We represent such mobility model with $RWP(v_{min}, v_{max}, t_p)$. It is worth nothing that epochs are not synchronized among different SUs.

The primary network consists of N_p PUs that are distributed in the network uniformly at random and each PU operates in some of the M channel or subbands. As in [118], to model each PU's activity at each channel, we adopt an independent two-state Markov chain alternating between the states busy (B) and empty (E). Let $\alpha_{l,i}$ and $\beta_{l,i}$ be the probabilities that the channel i of PU l switches its state from B to E and from E to B , respectively, for all $i = 1, 2, \dots, M$ and $l = 1, 2, \dots, N_p$. The utilization of channel i of PU l is given by $\lambda_{l,i} = \frac{\beta_{l,i}}{\beta_{l,i} + \alpha_{l,i}}$ [118].

The Frame Structure of CRNs

The SUs are assumed to be synchronized and operate in time on a frame-by-frame structure as in [118, 119]. The frame structure of a CRN, as shown in Figure 6.1, includes a sensing time T_S and a transmission time T_X that add up to the total frame time T . During T_S all SUs cease their transmission, perform spectrum sensing, and report the sensing results on a dedicated common control channel to the BS. As depicted in Figure 6.1, the sensing time T_S is comprised of two parts, namely *channel sensing time* (T_C) and *overhead time* (T_O). During channel sensing time T_C , each SU senses a PU channel. During overhead time T_O , SUs report the sensing results to the BS. Moreover, the BS assigns each SU a channel to sense in the next frame, and performs spectrum access assignment. In standard IEEE 802.22, T has been set at about a few hundred milliseconds [118, 120].

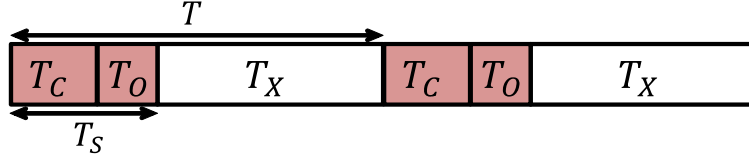


Figure 6.1: The frame structure of an SU's operation in a CRN depicting two consecutive time frames. During the sensing time T_S , all SUs cease their transmissions.

SU's Belief Vector

Because of the geographically dispersed SU network, PUs transmission can only be detected within a specific area. Outside that area, an SU can use the same channel for its transmissions (*frequency reuse*). This implies that PU channels' availability information is *inconsistent* in geographically dispersed SUs. Due to limited sensing capability of the SUs the state of every PU at every SU location cannot be observed. However, each SU may infer the state of PUs from the observation history. To this aim, let us define the *belief vector* $\mathbf{x}_j(t) \triangleq [x_{j,1}(t), \dots, x_{j,M}(t)]$, where $x_{j,i}(t)$ is the probability that SU j finds channel i empty at time t (whether or not it actually senses it). Let $a_j(t)$ denote the channel that SU j senses at time t . Moreover, let $S_{a_j(t)}(t) \in \{B, E\}$ be the status of the observed channel by SU j at time t .

The belief vector for SU j at time $t + 1$ is found as follows

$$x_{j,i}(t+1) = \begin{cases} 1 - \bar{\beta}_{j,i}(t), & a_j(t) = i, S_{a_j(t)} = E \\ \bar{\alpha}_{j,i}(t), & a_j(t) = i, S_{a_j(t)} = B \\ x_{j,i}(t)(1 - \bar{\beta}_{j,i}(t)) \\ + (1 - x_{j,i}(t))\bar{\alpha}_{j,i}(t), & a_j(t) \neq i. \end{cases} \quad (6.1)$$

In (6.1), $\bar{\alpha}_{j,i}(t)$ and $\bar{\beta}_{j,i}(t)$ are the state transition probabilities of channel i from the perspective of SU j at time t . While $\alpha_{l,i}$ and $\beta_{l,i}$ are defined for every PU, $\bar{\alpha}_{j,i}$ and $\bar{\beta}_{j,i}$ are defined from SUs' standpoint. Therefore, due to mobility of SUs, $\bar{\alpha}_{j,i}(t)$ and $\bar{\beta}_{j,i}(t)$ are constantly changing.

In addition, finding $\bar{\alpha}_{j,i}(t)$ and $\bar{\beta}_{j,i}(t)$ in terms of $\alpha_{l,i}$ and $\beta_{l,i}$, respectively, is not possible due to the lack of location information. After each SU completed the sensing at time t , it transmits the sensing decision to the BS and then BS calculates the beliefs based on (6.1) and stores $\mathbf{x}_j(t+1)$ for all $j = 1, 2, \dots, N$. Consequently, the BS determines the sensing policy $\mathbf{a}(t+1) \triangleq [a_1(t+1), \dots, a_N(t+1)]$, where $a_j(t+1)$ determines the channel that SU j senses at time $t+1$.

Cluster-CMSS: Cluster-Based Coordinated Multiband Spectrum Sensing

In this section, we explain our proposed policy, referred to as Cluster-CMSS, to find the optimal sensing policy. First we consider a scenario where the PU's dynamic (i.e., $\alpha_{l,i}$ and $\beta_{l,i}$ for all $i = 1, 2, \dots, M$ and $l = 1, 2, \dots, N_p$) is known. Later in this section, we consider a scenario where the dynamic of the PU activity is learned.

The Cluster-CMSS Policy With Known PU Dynamic

When $\bar{\alpha}_{j,i}(t)$ and $\bar{\beta}_{j,i}(t)$ for all $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$ are known, the BS can easily update the belief vectors using (6.1). The Cluster-CMSS algorithm is initialized to $\mathbf{x}_j(1) = [\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}]^T$ for all $j = 1, 2, \dots, N$. Therefore, the BS assigns a channel to each SU uniformly at random. At the beginning of the consequent time frames, the BS, after receiving the sensing results from the SUs, performs the following steps. The BS updates the belief vectors for all SUs, based on which SUs are partitioned into several clusters. For every cluster, the BS determines the unique channels to be sensed in the next time frame by performing a one-to-one matching algorithm between the members of that cluster and the channels. Algorithm 2 represents the pseudo code of the steps taken at the beginning of each frame. The time required to perform Algorithm 2 at each frame is equal to T_S , in which Step 1 takes T_C seconds and all other steps together take T_O

seconds.

Algorithm 2 The proposed Cluster-CMSS algorithm at the beginning of frame t

- 1: SUs sense the assigned channels.
 - 2: The BS receives the sensing results (B or E) from SUs.
 - 3: The BS determines the belief vectors $\mathbf{x}_j(t+1)$ for all $j = 1, 2, \dots, N$ (Eq (6.1)).
 - 4: The BS partitions SUs into clusters.
 - 5: The BS performs bipartite matching within each cluster and assigns each SU a channel to sense in the next time frame (some SUs will remain inactive).
 - 6: The BS transmits the channel access permissions and the ID of the channel that each SU has to sense at frame $t+1$.
-

The flowchart of our proposed policy is given in Figure 6.2. The tasks during T_S at the BS and SUs are depicted in right and left boxes, respectively.

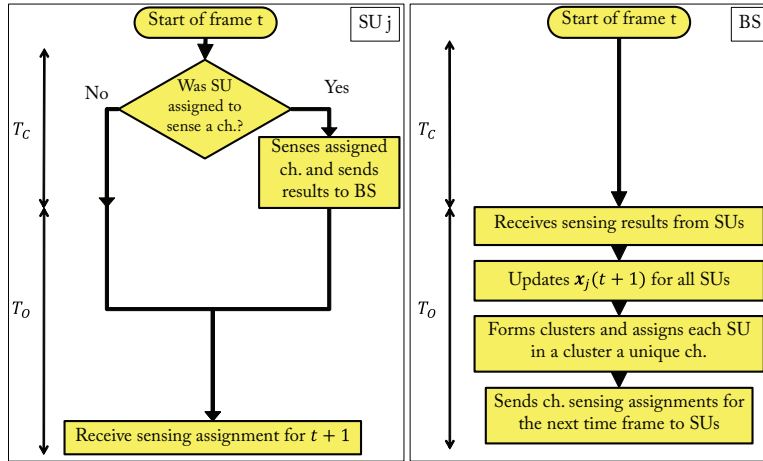


Figure 6.2: The flowcharts of the proposed Cluster-CMSS policy. The tasks during T_S at the BS and SUs are depicted in right and left boxes, respectively.

To measure the overall performance of the proposed policy, we define average *spectrum opportunity discovery ratio*, \bar{R}_s . This is the ratio of the average number of *unique* spectrum holes discovered per time frame, \bar{n}_u , to the total number of sensing attempts per time frame, N . This can be evaluated by averaging the instantaneous ratio of these parameters over time (i.e., $\bar{n}_u = E[n_u(t)]$).

At time t , the number of unique spectrum holes can be found by subtracting the number of duplicate sensed spectrum holes, $n_d(t)$, from the total number of successful sensing attempts, $n_s(t)$. If two or more SUs are located within transmission range of each other and they sense the same channel empty in one time frame, one of these sensing attempts is considered unique and the rest are duplicate spectrum holes. Therefore, \bar{R}_s is found as follows

$$\bar{R}_s = \frac{\bar{n}_u}{N} = \frac{\bar{n}_s - \bar{n}_d}{N} = \frac{E[n_s(t)] - E[n_d(t)]}{N}. \quad (6.2)$$

In Figure 6.3, we provide an example of a spectrum sensing assignment. In this example $N = 6$, $M = 3$, and we have 3 clusters each containing two SUs (see Figure 6.3(a)). Figure 6.3(b) represents the channel sensing assignments on a bipartite graph between the SUs (square nodes) and the PU channels (circular nodes) as well as the state of the PU channels (B and E stand for the busy and the empty states, respectively). The details of the spectrum sensing assignment will be discussed in the next section. Using the assignment represented by the edges of bipartite graph in Figure 6.3(b) at time t , the SUs $\{1, 2, 3, 5, 6\}$ will sense an empty channel. However, given SUs 1 and 3 are within the range of each other their sensing of channel 1 finds a duplicate spectrum hole. Therefore in this example, $n_s(t) = 5$, and $n_d(t) = 1$. Hence, we find the spectrum opportunity discovery ratio at time t as $R_s(t) = \frac{5-1}{6} = 0.67$. In the rest of this section, we separately describe the building blocks of our proposed Cluster-CMSS policy.

Learning the PU Dynamic

In most real-world scenarios the dynamic of the PUs arrival and departure are a priori unknown. Therefore, the BS needs to learn them on the fly. A simple and practical method of learning PU's dynamic is by recording the sample means of $\bar{\alpha}_{j,i}(t)$ and $\bar{\beta}_{j,i}(t)$ for all $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$ [121]. In other words, the BS station determines the number of times that each SU observes a certain channel has changed its state from empty to busy and vice versa. Therefore,

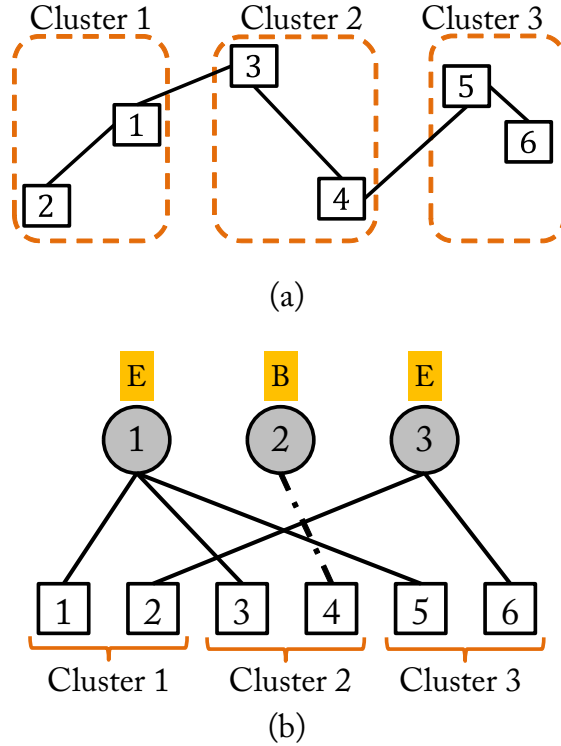


Figure 6.3: Channel sensing assignment for a network with $N = 6$ (SUs are represented by squares) and $M = 3$ (channels are represented by circles). Subfigure (a) represents the location of SUs. There is a line between two SUs if they are in the transmission range of each other. Subfigure (b) represents the channel sensing assignments and the state of each PU channel on a bipartite graph. In addition the edges with dash-dotted line and solid line represent the failed and the successful sensing attempts, respectively.

the estimated values of these parameters, $\hat{\alpha}_{j,i}(t)$ and $\hat{\beta}_{j,i}(t)$, will be used in (6.1). To estimate these parameters, we define a *learning window* with length T_l time frames in which the number of state transitions is counted. When the SUs are static, increasing the length of the learning window will add to the accuracy of the parameter estimations. However, in mobile SUs scenarios having a lengthy learning window reduces the accuracy of parameter estimation because of SUs movements. In the following, we determine the optimal length of learning window under random waypoint mobility model for SUs.

We propose to choose the length of the learning window equal to the average time it takes a mobile SU to move out of an active PU's range. In other words, the previous sensing results of an SU that are older than this average time are no longer useful in determining the PU's dynamic. The following theorem provides tight upper and lower bounds for the average time it takes for an SU to move out of an active PU's range.

Theorem 2. *Assume a circle with radius R is entirely located at random inside an area A with a rectangular shape. Given an SU exists within the boundaries of this circle. The average time it takes for this SU, which moves based on random waypoint model $RWP(v_{min}, v_{max}, t_p)$, remains inside that circle is denoted by T_r and is bounded as follows,*

$$\frac{2}{v_{min} + v_{max}} \left(\frac{P_{in}}{1 - P_{in}} \frac{128R}{45\pi} + 1 \right) \leq T_r \leq \frac{2}{v_{min} + v_{max}} \left(\frac{P_{in}}{1 - P_{in}} \frac{128R}{45\pi} + \frac{4}{3} \right), \quad (6.3)$$

where P_{in} is the probability that a waypoint falls inside the circle and is given by $P_{in} = \frac{\pi R^2}{A}$

Proof. To prove Theorem 2, assume an SU is located inside a circle with radius R , we want to find the average amount of time that it takes for the SU to leave that circle. The mobility model is $RWP(v_{min}, v_{max}, t_p)$. Hence, the SU will leave the circle if its new waypoint lies outside of the circle with radius R . Let P_{in} be the probability that a waypoint is chosen inside the circle with radius R and is denoted by $P_{in} = \frac{\pi R^2}{A}$. The number of epochs that it takes the SU to leave the circle follows a geometric distribution with success probability $1 - P_{in}$. Accordingly on average it will take $\frac{P_{in}}{1 - P_{in}}$ epochs before it leaves the circle. Given waypoint is located inside the circle. The average time of an epoch with a waypoint inside the circle R , is T_{in} , and found as follows

$$T_{in} = \frac{\frac{128R}{45\pi}}{\frac{v_{min} + v_{max}}{2}}. \quad (6.4)$$

in (6.4), the numerator is the average distance between any two points in a circle with radius R chosen uniformly at random [122] and the denominator is the average velocity. On the other hand

given the waypoint falls outside of the circle, it will take T_{edge} seconds on average until it reaches the edge of the circle and leaves it. Accordingly T_{edge} is determined by finding the average distance of random point inside a circle to any point in its circumference and dividing it to the average speed. Within a circle with radius R located at the origin, the average distance of a point at location $(r, 0)$ to any point in its circumference is found as follows

$$\begin{aligned} L_{edge}(r) &= \frac{1}{2\pi} \int_0^{2\pi} \sqrt{(R \cos \phi - r)^2 + R^2 \sin^2 \phi} d\phi \\ &= \frac{2}{\pi} \left| 1 - \frac{r}{R} \right| E_e \left(-\frac{4r/R}{(1 - r/R)^2} \right), \end{aligned} \quad (6.5)$$

where $E_e(r)$ is the complete elliptical integral of the second kind, which is defined as $E_e(r) \triangleq \int_0^{\frac{\pi}{2}} \sqrt{1 - r^2 \sin^2(\theta)} d\theta$. Therefore, we find T_{edge} as follows,

$$T_{edge}(r) = \frac{L_{edge}}{\frac{v_{min} + v_{max}}{2}} = \frac{\frac{4}{\pi} \left| 1 - \frac{r}{R} \right|}{v_{min} + v_{max}} E_e \left(-\frac{4r/R}{(1 - r/R)^2} \right). \quad (6.6)$$

It is easy to verify that for all $0 \leq \frac{r}{R} \leq 1$, we have $1 \leq \frac{2}{\pi} \left| 1 - \frac{r}{R} \right| E_e \left(-\frac{4r/R}{(1 - r/R)^2} \right) \leq \frac{4}{3}$. Therefore the value of $1 \leq T_{edge} \leq \frac{4}{3}$.

Accordingly the average time it takes an SU to move out of the range of a PU is

$$T_r = \frac{P_{in}}{1 - P_{in}} T_{in} + T_{edge} \quad (6.7)$$

The upper bound and the lower bound on (6.7) can be found as follows

$$\frac{2}{v_{min} + v_{max}} \left(\frac{P_{in}}{1 - P_{in}} \frac{128R}{45\pi} + 1 \right) \leq T_r \leq \frac{2}{(v_{min} + v_{max})} \left(\frac{P_{in}}{1 - P_{in}} \frac{128R}{45\pi} + \frac{4}{3} \right). \quad (6.8)$$

□

Using Theorem 2, we set the length of the learning window to be the closest multiple of T to the midpoint of the upper and lower bounds given in (6.3). That is

$$T_l = \text{round} \left(\frac{2}{(v_{min} + v_{max})T} \left(\frac{P_{in}}{1 - P_{in}} \frac{128R}{45\pi} + \frac{7}{6} \right) \right). \quad (6.9)$$

In (6.9), the function $\text{round}(\cdot)$ rounds its argument to the closest integer. Clearly when SUs are static ($v_{max} = v_{min} = 0$), both the upper and the lower bounds of T_l goes to ∞ that means the length of learning window can grow very large. In other words we can use all of previous sensing results to estimate the channel parameters. Given the length of the learning window, T_l , the BS can calculate the average number of state transitions that it has observed per channel for each SU during the learning window. For instance $\hat{\alpha}_{j,i}(t)$ is determined by dividing the number of times SU j observes channel i changed its state from B to E to the number of times the SU j has observed the state of channel i to be in B state during past T_l sensing attempts. The value of $\hat{\beta}_{j,i}(t)$ is calculated in the same way for all SUs and PU channels. It is worth noting that in some cases, specially when the velocity of SUs are high, an SU may not observe one or more channels during T_l . In those cases, the estimation of PU parameters from previous time frame will be used.

The Sensing-Based Clustering

In the geographically dispersed networks, clustering allows frequency reuse and more efficient spectrum sensing. By grouping the nodes that share a same set of spectrum holes, the BS can coordinate sensing assignment among members of every cluster. In the lack of SU's location information, we propose to use the *sensing results of SUs* as a *clustering metric*. We define the distance between two SUs based on the distance between their belief vectors. More specifically, we define the distance D_x between any two SUs as the Kullback-Leibler (KL) divergence between beliefs of those SUs. In other words, the distance is measured by the divergence in the beliefs of

SUs j_1 and SU j_2 and is defined as follows

$$D_x(j_1, j_2) \triangleq D_{KL}(\mathbf{x}_{j_1}(t) \parallel \mathbf{x}_{j_2}(t)) + D_{KL}(\mathbf{x}_{j_2}(t) \parallel \mathbf{x}_{j_1}(t)), \quad (6.10)$$

where $D_{KL}(\mathbf{x}_{j_1}(t) \parallel \mathbf{x}_{j_2}(t)) \triangleq \sum_{i=1}^m x_{j_1,i}(t) \log \frac{x_{j_1,i}(t)}{x_{j_2,i}(t)}$. If two SUs experience exactly the same set of observations on PUs' channels, they will have the same beliefs on PU's channels and the KL distance between them will be zero. Similarly, SUs with different PU's channel sensing experiences will have diverged beliefs and consequently greater distances.

Various clustering algorithms have been proposed in the literature for different purposes in CRNs. In our case, we are interested in a clustering algorithm that provides *hard partitioning*, has *low complexity*, and operates without the prior knowledge on PU's dynamic. To meet these requirements and to cluster SUs, we integrate our proposed *sensing distance* metric (defined in (6.10)) into the k -means clustering method. Since the number of clusters is not known a priori we use the *elbow method* to determine the number of clusters. Accordingly, we start with $k = 1$ cluster and find the within cluster point-to-centroid distance variance. By increasing k , within cluster point-to-centroid variance decreases. However, there exists a k , k_{opt} , beyond which increasing the number of cluster will only improve the variance marginally. This point is called the elbow point [123].

Coordinated Spectrum Sensing Within Clusters Using Bipartite Matching

In this section, we describe the mapping of the CMSS problem onto a bipartite matching problem. After the BS clusters the SUs, it assigns each SU within each cluster a unique channel to sense. The goal is to assign every SU to sense the channel it believes to have the highest probability of being empty. For each cluster, the BS solves this problem by finding a *minimum-weight matching* on a bipartite graph that is constructed as follows.

The vertices of one side of the graph correspond to the SUs in a cluster (i.e., N_k vertices) and the vertices of the other side of the graph correspond to the PU channels (i.e., M vertices). An edge exists between any two vertices from each side of this bipartite graph with a positive weight (See Figure 6.4). We inversely relate $w_{j,i}(t)$, the weight of the edge connecting SU j to channel i , to $x_{j,i}(t)$, the belief SU j on channel i . Therefore, the greater $x_{j,i}(t)$, the smaller the weight of edge between channel i and SU j would be. The weights of each edge is set as $w_{j,i}(t) = \frac{1}{x_{j,i}(t)+\epsilon}$, where ϵ is a very small constant to avoid unbounded weights. Using this strategy, we find the minimum-weight allocation which corresponds to maximizing the probability of finding an empty channel for each cluster member. We employ the well-known Hungarian Algorithm [124] to solve the minimum-weight matching problem. Algorithm 3 represents the proposed intra-clustering assignment. Figure 6.4 depicts an example of CMSS within a cluster using bipartite matching. In this example, $M = 3$ (circles), for this cluster $N_k = 2$ (squares), and the weight of each edge is represented by its corresponding edge. The double-lined edges represent the minimum weight matching and the dashed edges represent the unmatched edges. Based on this matching, SU 1 and SU 2 will sense channels 2 and 3, respectively.

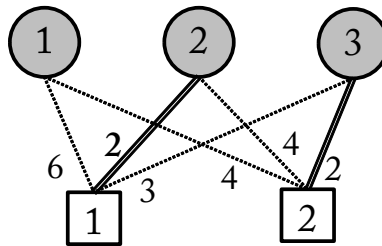


Figure 6.4: An example of CMSS within a cluster using bipartite matching.

Algorithm 3 The intra-cluster CMSS algorithm at time frame t .

- 1: **for** every cluster **do**
 - 2: Calculate the weights according to $w_{j,i}(t) = \frac{1}{x_{j,i}(t)+\epsilon}$ for all $j \in \{1, \dots, N_k\}$ and $i \in \{1, \dots, M\}$.
 - 3: Run the minimum-weight *Hungarian* algorithm [124].
 - 4: **end for**
 - 5: Transmit the obtained channel sensing assignment results to SUs.
-

Performance Evaluation of Cluster-CMSS

In this section, we study the performance of our proposed Cluster-CMSS algorithm. Suppose $A(t)$ is the set of SUs that has been assigned by the BS to perform spectrum sensing at time frame t and $|A(t)|$ be the cardinality of $A(t)$. We find the probabilities of miss-detection $P_m^j(t)$ and false-alarm $P_f^j(t)$ for all $j \in A(t)$ under the additive white Gaussian Noise (AWGN), Rayleigh, and Rician channel models. The average probabilities of miss-detection $Q_m(t)$ and false-alarm $Q_f(t)$ are given by,

$$Q_m(t) = \frac{1}{|A(t)|} \sum_{j \in A(t)} P_m^j(t), \quad (6.11)$$

$$Q_f(t) = \frac{1}{|A(t)|} \sum_{j \in A(t)} P_f^j(t). \quad (6.12)$$

The SUs that are not assigned to sense any channel do not contribute to the $Q_m(t)$ and $Q_f(t)$. In the following for the brevity of expressions we omit the variable t in all formulas.

Miss-Detection and False-Alarm probabilities over AWGN channels

Suppose γ_j is the received SNR at SU j . An exact closed-form expression for the probabilities of miss-detection P_m^j and false-alarm P_f^j of SU j over the AWGN channel are as follows [125],

$$P_f^j = \frac{\Gamma(T_C W, \frac{\delta}{2})}{\Gamma(T_C W)}, \quad (6.13)$$

$$P_m^j = 1 - Q_{T_C W}(\sqrt{2\gamma_j}, \sqrt{\delta}), \quad (6.14)$$

where δ is the decision threshold, $\Gamma(\cdot)$ is the gamma function, $\Gamma(\cdot, \cdot)$ is the incomplete gamma function, and $Q_{\square}(\cdot, \cdot)$ is the generalized Marcum Q-function [125]. Without loss of generality, we choose the value of T_C such that $T_C W$ is restricted to be an integer. The value of T_C can be determined such that it keeps P_f^j and P_m^j below predefined thresholds for all $j \in \{1, \dots, N\}$.

Miss-Detection and False-Alarm Probabilities Over Rayleigh and Rician Channels

In many practical networks, the spectrum sensing quality might be adversely affected by fading. In this section, we briefly consider the scenarios in which the SNR of the sensed signal at SUs follows Rician and Rayleigh distributions. Rician model represents the scenarios in which SUs receive the PU signal from several different paths, with one direct path that is stronger than the others. Rician factor K is the ratio between the power received from the direct path and the power received from other scattered paths [126]. Rayleigh model is suitable for scenarios where the direct path does not exist. Therefore, the Rayleigh fading channel is a special case of Rician fading channel with $K = 0$. Clearly, P_f^j remains the same under the fading scenario because P_f^j concerns the case of no signal transmission and hence is independent of the received SNR. In case of Rician channel the received SNR γ_j is a random variable that follows Rician distribution. The PDF of γ_j (in dB) for all $\gamma_j > 0$ is given by [125],

$$f(\gamma_j) = \frac{K+1}{\bar{\gamma}_j} \exp\left(-K - \frac{(K+1)\gamma_j}{\bar{\gamma}_j}\right) I_0\left(2\sqrt{\frac{K(K+1)\gamma_j}{\bar{\gamma}_j}}\right), \quad (6.15)$$

where $\bar{\gamma}_j$ is the average SNR (in dB) at SU j and can be estimated as described in [126], and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind.

The probability of miss-detection P_m^j can be obtained by averaging (6.14) over the Rician distri-

bution in (6.15). A closed-form expression is given in [125] for special case of $T_C W = 1$,

$$P_m^j|_{T_C W=1} = Q_{T_C W=1} \left(\sqrt{\frac{2K\bar{\gamma}_j}{K+1+\bar{\gamma}_j}}, \sqrt{\frac{\delta(K+1)}{K+1+\bar{\gamma}_j}} \right). \quad (6.16)$$

For $K = 0$ this expression reduces to Rayleigh fading scenario [125]. Similar to the AWGN scenario, the average probabilities of miss-detection and false-alarm are found using (6.11) and (6.12).

Energy Cost of Sensing

One of the important concerns in the design of the CRNs is the energy cost of the spectrum sensing because it is a major contributor to the total energy consumption. Suppose the energy cost of sensing one channel by an SU is E_S and the energy costs associated with reporting the sensing results are E_{TX} (corresponding to transmitter energy consumption at the SU) and E_{RX} (corresponding to receiver energy consumption at the BS). The energy costs associated with BS informing an SU of the channel to sense in the next frame are E_{TX} for the BS and E_{RX} for the SU. In addition, The energy cost of idling during T_S is E_{id} .

The energy costs associated to sensing depending on whether or not an SU is assigned to sense a channel are E_1 and E_2 , respectively and given by

$$E_1 = E_S + 2(E_{TX} + E_{RX}), \quad (6.17a)$$

$$E_2 = E_{id} + E_{TX} + E_{RX}. \quad (6.17b)$$

In (6.17a), an SU has to report its sensing results to the BS and consequently the BS sends that SU the information about the sensing assignments for the time frame. Therefore, the cost of communication with the BS (i.e., $E_{TX} + E_{RX}$) is included twice.

In comparison, the energy cost of spectrum sensing in the greedy non-cooperative policy [5], E_t^g , is found as $E_t^g = E_S + E_{TX} + E_{RX}$, noting that all the SUs independently choose to sense the best possible channel and transmit the outcome of sensing that channel to the BS. Similarly, the energy cost of spectrum sensing in the genie-aided spectrum sensing policy, E_t^{ga} , is found as $E_t^{ga} = ME_S + E_{TX} + E_{RX}$. In this case, all the SUs sense the entire spectrum band and transmits the results to the BS. In our numerical simulations, we compare energy cost of our proposed policy with these two policies assuming the cost of accessing an empty channel is equal to E_{acc} for all three policies. The energy cost per successful SU channel access is the sum of average energy cost to find an empty channel plus the cost to access a channel (i.e. E_{acc}). z

Numerical Results

For our numerical simulations, we set $N = 50$, $M = 10$, $N_p = 20$ and assume SUs are distributed uniformly at random in an area with size $A = 1000^2$ (distance unit)². At this point for all PUs, we assume the AWGN channel scenario and the received signal power is only affected by path loss with path loss exponent $\gamma = 2.7$. We suppose SUs can detect each PU's transmission within 100 distance unit range with a high probability. For the brevity of the results, we assume the channels for all PUs have similar parameters (i.e., $\alpha_{l,i} = \alpha$, $\beta_{l,i} = \beta$, and $\lambda_{l,i} = \lambda$ for all $i = 1, \dots, M$ and $l = 1, \dots, N_p$). We set $\alpha = 0.1$ and change the value of β to obtain the desired channel utilization λ . For Cluster-CMSS, the number of clusters is determined using the elbow method as previously described. The results of this simulation is shown in Figure 6.5, which represents the *average spectrum opportunity discovery ratio* (\bar{R}_s), as defined in (6.2), versus λ . In addition, the SUs move according to the random waypoint model $RWP(0, 15, 2.5)$.

In Figure 6.5, we compare the spectrum sensing performance of our proposed Cluster-CMSS policy with a genie-aided location aware policy and the greedy non-cooperative spectrum sensing

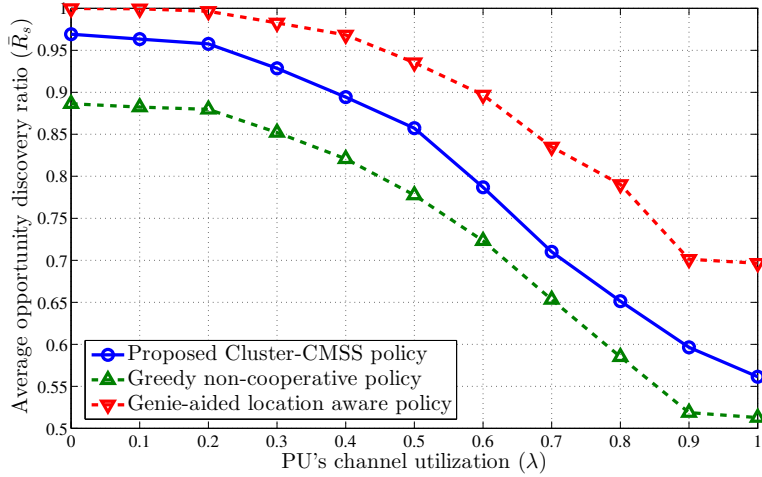


Figure 6.5: Average spectrum opportunity discovery ratio versus the PU's channel utilization λ .

policy in [5]. In the genie-aided sensing policy, the BS is aware of the status of the previous channel states at all SUs and the distance between SUs. Clearly implementing the genie-aided policy in a geographically dispersed and mobile network is impractical. Therefore, the genie-aided policy solely serves as a performance upper bound. As we can see, when PU channels are under-utilized all policies have a high opportunity discovery rate due to abundance of spectrum holes. However, when λ is close to 1 (heavy PU utilization) our proposed policy performs better than the greedy non-cooperative policy by at least 15%. By increasing the channel utilization, the spectrum holes become more scarce and the effectiveness of the proposed policy in finding spectrum holes becomes more lucid.

In Figure 6.6, we have depicted the average opportunity discovery ratio of Cluster-CMSS versus the maximum velocity of the SU in the random waypoint model. We set the same parameters as previous simulation ($N = 50$, $M = 10$, $N_p = 20$, $\lambda = 0.5$ and $\gamma = 2.7$). The mobility model in this simulation is $RWP(0, v_{max}, 2.5)$. As we can see the average opportunity discovery ratio decreases by increasing v_{max} . According to (6.9), by increasing v_{max} the length of the learning

window decreases. Smaller learning window reduces BS's capability to learn the dynamic behavior of the PU network. Hence, Cluster-CMSS will not be able to effectively employ the information from previous channel occupancy of PUs.

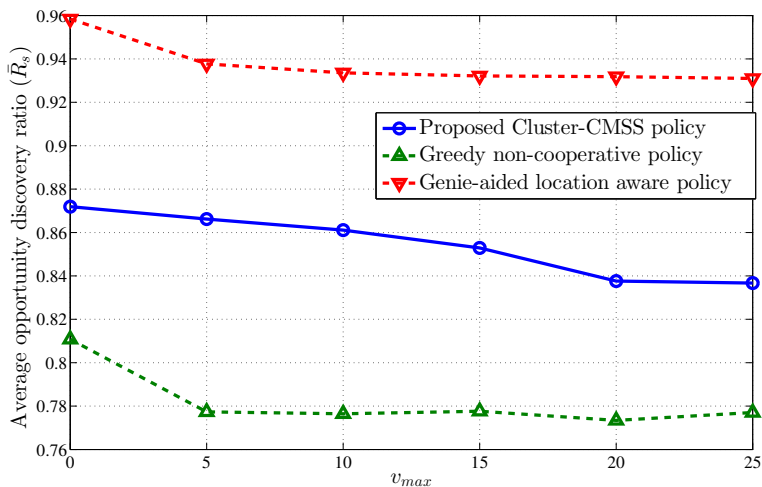


Figure 6.6: Average spectrum opportunity discovery ratio versus maximum velocity of SU's movement (v_{max}).

In Figure 6.7, we have depicted the characteristic graph (the probability of miss-detection versus the probability of false-alarm) of Cluster-CMSS. The simulation parameters are similar to the previous simulations and we have considered the effects of non-ideal sensing. The analytical results obtained using (6.13) and (6.14) are compared against the values obtained through the numerical simulations. Clearly, by increasing the received SNR at each SU, the overall probabilities of miss-detection and false-alarm decrease. In addition, the shaded area in Figure 6.7 represents the area in which the values of Q_m , Q_f , or both are not acceptable by the IEEE 802.22 standard [127]. Therefore, it is important to have the values of both Q_m and Q_f less than or equal to 0.1.

In Figure 6.8, we represent the numerically-obtained characteristic graph of Cluster-CMSS and the non-cooperative policy in [5], under AWGN, Rician, and Rayleigh channels assuming the

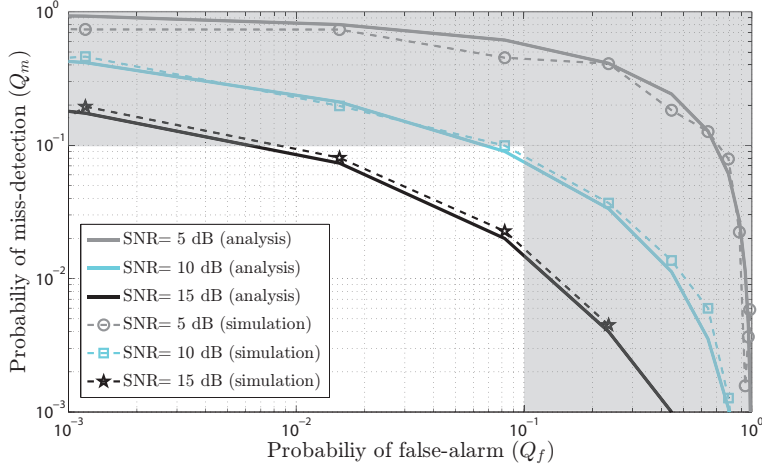


Figure 6.7: The characteristic graph (the probability of miss-detection versus the probability of false-alarm) of Cluster-CMSS. Each curve represents different average SNR at the sensing SUs. The performance in the shaded areas are not allowed in the IEEE 802.22 standard.

average received SNR is 15 dB. In this simulation, we employ the same simulation parameters as in previous experiment, and the Rician factor is $K = 10$. As we can see, Cluster-CMSS has a better performance under all channel conditions compared to non-cooperative policy [5]. Moreover, Cluster-CMSS is the most effective in the AWGN scenario.

Now let us pair Cluster-CMSS with a very simple spectrum access scheme, which basically allows every SU to access the channel it finds empty and transmit on that channel. The simulation parameters are similar to previous simulations (i.e., $N = 50$, $M = 10$, $N_p = 20$, $\lambda = 0.5$ and $\gamma = 2.7$). At Table 6.1, we compare the average energy costs per successful SU transmission in one time frame for different policies. As reported in [128], the energy cost of an SU to sense one channel, to transmit/receive a channel sensing result, to access a channel, and to idle during sensing time is $E_S = 3.5mJ$, $E_{TX} = E_{RX} = 0.1125mJ$, $E_{acc} = 4mJ$, and $E_{id} = 0.05mJ$, respectively. Therefore, we find the energy cost of different policies using the simple access scheme (every SU accesses the channel it finds empty).

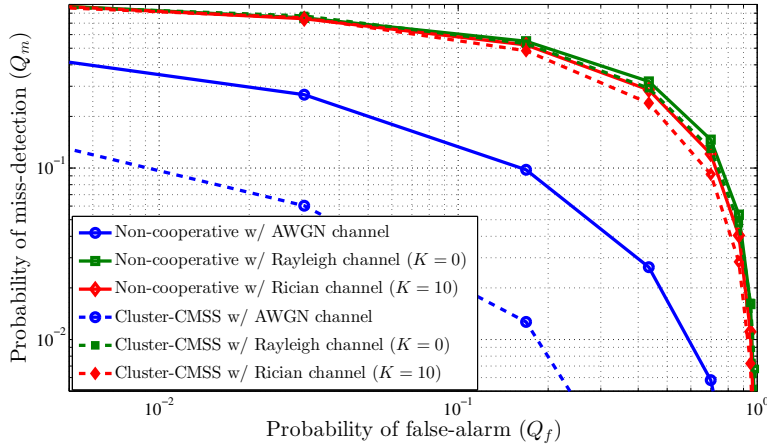


Figure 6.8: The numerically obtained characteristic graph of the proposed Cluster-CMSS policy and the non-cooperative greedy policy in [5] at 15 dB average received SNR.

Table 6.1: Average energy cost per successful SU transmission

Spectrum sensing policy	Energy cost (mJ)
Cluster-CMSS	7.97
Greedy non-cooperative	7.725
Genie-aided	39.22

As it can be concluded from Table 6.1, the energy cost of Cluster-CMSS is slightly higher, because of coordination overhead, than non-cooperative greedy policy. This slight increase in energy consumption is the price of larger opportunity discovery ratio in the spectrum sensing. In addition, we have included the energy cost that it is required to implement the genie-aided policy, which is considerably larger than our proposed Cluster-CMSS and greedy policies.

Application of Structured Sparse Compressed Sensing in Cognitive Radio Spectrum Sensing

Authors in [19–21] have considered the problem of spectrum sensing and PU localization in cognitive radio networks. They used compressive sensing to recover the location and the transmission power of PUs on a grid of potential points. In [129] the authors have solved the problem of PU localization using Bayesian compressive sensing algorithm without including the structure in the data.

Here we adopt a system model similar to what is assumed in [19]. Suppose an area of interest is divided into N_s grid points (as shown in Figure 6.9) and transmitters (PUs) are assumed to be located in a subset of these grid points, unknown to us. On the other hand there are N_r receivers (SUs) with known locations that receive a superposition of the transmitters' signals affected by the channel gains γ_{sr} between the transmitters and receivers and observed in the presence of zero mean AWGN with variance σ_r^2 . The goal is to estimate the power spectral density (PSD) at each point, based on which we can localize the PUs and find the frequency bands that they have occupied.

While a solution to above problem could be found through exhaustive search with a very high computational complexity, exploiting the inherent *sparsity* in the problem can make the problem tractable. Sparsity comes from scarce presence of active transmitters as well as narrow-band transmissions. Authors in [19] formulated this following a Lasso problem.

The model of PUs being located at possible grid points not only imposes the sparsity but also provides some extra information on the structure of sparse signal. In the primary network, mostly due to interference avoiding considerations, PUs will not be present in the neighboring grid locations. In other words, if a PU is found active at one location, there exist a neighborhood around that PU such that no other PU does not exist in that neighborhood (matern hardcore model [130]). The minimum distance between PUs are determined by their transmit power. However, for simplicity

we assume all PUs transmit with the same power. Therefore, all PUs will have similar range. For example the structure of the PUs' grid is represented in Figure 6.9, where existence of a PU at a grid point indicates that none of its neighboring grid points is occupied by other PUs. In this particular example the range of each PU includes the four immediate neighbors of that grid point.

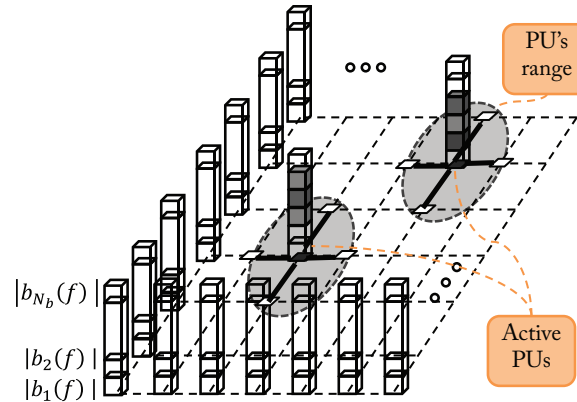


Figure 6.9: The grid of the N_s PU candidate locations. The PUs are present in 2 locations and no other PU is present in the range of those PUs. Therefore the white grid point does not include any PU. Each column represents the activity of PUs in different frequency channel

Accordingly, the problem of localization and estimating the PSD of PUs in a cognitive radio network could be solved using a structured sparse CS recovery.

Conclusion

In this chapter, we considered the problem of coordinated multiband spectrum sensing in the geographically dispersed and mobile cognitive radio networks. We proposed a policy that detects the spectrum holes without depending on the location information of the primary users. According to our proposed policy, the secondary users are clustered based on their spectrum sensing results.

We introduced a novel metric for clustering SU nodes, which is based on the consensus among the SUs' channel sensing results. In our proposed policy, the BS uses this metric to form the clusters without the need to know the location of the SUs. Then, the BS performs a graph-theory-based coordinated spectrum sensing among members of each cluster. For the mobile SUs that move according to a random way point model, we have shown through extensive simulations that the proposed policy considerably increases the spectrum opportunity discovery ratio for the secondary users at the cost of slight increase in the energy consumption associated with spectrum sensing.

CHAPTER 7: CONCLUSION

In this dissertation, we have investigated the theoretical and practical challenges in design and implementation of the signal processing algorithms that are exploiting the underlying structured sparsity model, which exists in many real-world signals. Examples of signals with structured sparsity are natural images, readings of wireless sensor networks, and electromagnetic spectrum occupation in spatial and temporal domains.

First, we designed a compressive sensing recovery algorithm for time-correlated sparse signals and investigated its performance under model mismatch scenarios. Next, we examined the problem of non-uniform sampling and recovery for the signals with non-uniform importance among signal coefficients. We showed that employing our proposed algorithm provides higher recovery accuracy for more important signal coefficients. Accordingly, employing the similar ideas using the properties of wavelet coefficients of images, we developed a model aware compressive sensing sampling and recovery algorithm for natural images that performs better than state-of-the art methods. Finally, we consider the problem of spectrum sensing in disperse cognitive radio networks and show that this problem can be formulated using a structured sparse CS recovery problem. In the rest of this Chapter, we summarize the contribution of this dissertation:

Time-Correlated Compressed Sensing

In Chapter 3, we proposed TC-CSBP, which is a CS recovery algorithm for sparse signals that are also time-correlated. TC-CSBP is based on belief propagation (CSBP) by Baron et al. CSBP serves as the underlying recovery scheme. However, we modify CSBP such that the priori knowledge about the signals coefficients time correlation are included in the algorithm. Our results show a

considerable improvement over conventional CSBP and other related work. Moreover, our results show that TC-CSBP is robust to the error in time-correlation model parameters to a great extent and it can maintain its supremacy in the presence of model mismatch.

Non-Uniform Compressed Sensing

In Chapter 4, we address the problem of non-uniform sampling and recovery of sparse signals. We propose NCS for non-uniform sampling of sparse signals. To realize NCS, we employ a *non-uniformly sparse* measurement matrix such that more important coefficients are captured by a relatively larger number of measurements. We analyzed the RIP of the measurement matrix of NCS. NCS can also be integrated with a recovery algorithm that exploits the non-uniform sparsity of signals [1, 3] for further performance improvement.

Non-Uniform Sampling and Recovery of Natural Images Using the Hidden Markov Tree Structure of Wavelet Coefficients

In Chapter *CSBP-HMT*, some of the ideas from previous two chapters are combined and are applied to the problem of sampling and recovery of natural images. Accordingly, we modify two Bayesian CS recovery algorithms to incorporate the signal structure (i.e. the Hidden Markov Tree structure of the wavelet coefficients). In addition based the underlying structure, we propose a measurement matrix that is designed incorporating the signal structure. Therefore, we develop a *model-based CS non-uniform sampling and recovery* scheme (uHMT-NCS) for natural images. We show, not only the signal model can be utilized to optimize the initial priors for Bayesian CS recovery algorithms, but it can also be employed in the design of new CS measurement matrices. The results of our numerical experiments suggest a significant performance gain compared to the

state-of-the-art model-based CS algorithms.

Coordinated Spectrum Sensing in Cognitive Radio Networks

The problem that we tackled in Chapter 6 is completely different than that of previous chapters. Here, the goal is to develop a spectrum sensing assignment policy that maximizes the discovery ratio of spectrum holes while the overhead spectrum sensing time is minimized by sensing only one channel per SU at a time. We propose *cluster-based coordinated multiband spectrum sensing* (Cluster-CMSS). In this problem SUs are mobile and can communicate with a central node or *base station* (BS). This is a very complex problem with numerous challenges. The main challenges are limited ability of SUs in sensing the spectrum, geographically dispersed SU distribution, dynamic PU activity and inaccurate sensing.

In cognitive radio networks, to construct the REM, the power spectral density (PSD) information of every point in space at each frequency must be available in real-time. Building on the spectrum sensing problem that is introduced in this chapter 6, we introduced a CS-based spectrum sensing and source localization scheme that employs the structure of the signal sparsity.

LIST OF REFERENCES

- [1] M. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, “Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 1985–2001, May 2011.
- [2] S. Som and P. Schniter, “Compressive imaging using approximate message passing and a markov-tree prior,” *IEEE Trans. Signal Process.*, vol. 60, pp. 3439–3448, Jul 2012.
- [3] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Tran. Inf. Theory*, vol. 56, pp. 1982–2001, Apr 2010.
- [4] L. He, H. Chen, and L. Carin, “Tree-structured compressive sensing with variational bayesian analysis,” *IEEE Signal Process. Lett.*, vol. 17, pp. 233–236, Mar 2010.
- [5] C. Tekin, S. Hong, and W. Stark, “Enhancing cognitive radio dynamic spectrum sensing through adaptive learning,” in *IEEE Military Communications Conference, 2009. MILCOM 2009.*, pp. 1–7, Oct 2009.
- [6] V. Turner, J. F. Gantz, D. Reinsel, and S. Minton, “The digital universe of opportunities: Rich data and the increasing value of the internet of things.” White Paper.
- [7] D. Baron, S. Sarvotham, and R. Baraniuk, “Bayesian compressive sensing via belief propagation,” *IEEE Trans. Signal Process.*, vol. 58, pp. 269–280, Jan 2010.
- [8] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey,” *Computer Networks*, vol. 50, pp. 2127–2159, May 2006.
- [9] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 201–220, Feb. 2005.

- [10] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, “Cooperative spectrum sensing in cognitive radio networks: A survey,” *Physical Communication*, vol. 4, pp. 40 – 62, Dec 2011.
- [11] Z. Quan, S. Cui, A. Sayed, and H. Poor, “Optimal multiband joint detection for spectrum sensing in cognitive radio networks,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 1128–1140, March 2009.
- [12] W.-Y. Lee and I. Akyildiz, “Optimal spectrum sensing framework for cognitive radio networks,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 3845 –3857, Oct 2008.
- [13] Y. Zhao, L. Morales, J. Gaeddert, K. Bae, J.-S. Um, and J. Reed, “Applying radio environment maps to cognitive wireless regional area networks,” in *New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on*, pp. 115–118, April 2007.
- [14] A. Vizziello, I. F. Akyildiz, R. Agusti, L. Favalli, and P. Savazzi, “Cognitive radio resource management exploiting heterogeneous primary users,” in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, IEEE, 2011.
- [15] H. Yilmaz, T. Tugcu, F. Alagoz, and S. Bayhan, “Radio environment map as enabler for practical cognitive radio networks,” *Communications Magazine, IEEE*, vol. 51, pp. 162–169, December 2013.
- [16] G. Mao, B. Fidan, and B. D. Anderson, “Wireless sensor network localization techniques,” *Computer Networks*, vol. 51, no. 10, pp. 2529 – 2553, 2007.
- [17] Z. Tian and G. Giannakis, “Compressed sensing for wideband cognitive radios,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, vol. 4, 2007.

- [18] Y. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 2337–2340, April 2009.
- [19] J. Bazerque and G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing*, vol. 58, pp. 1847–1862, March 2010.
- [20] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, pp. 2002–2016, May 2011.
- [21] J. Bazerque, G. Mateos, and G. Giannakis, "Group-lasso on splines for spectrum cartography," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4648–4663, Oct 2011.
- [22] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
- [23] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [24] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
- [25] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb 2006.
- [26] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

- [27] M. Davies and R. Gribonval, “Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p \leq 1$,” *Information Theory, IEEE Transactions on*, vol. 55, pp. 2203–2214, May 2009.
- [28] J. D. Blanchard, C. Cartis, and J. Tanner, “Compressed sensing: How sharp is the restricted isometry property?,” *SIAM Rev.*, vol. 53, pp. 105–125, Feb. 2011.
- [29] N. Vaswani and W. Lu, “Modified-CS: Modifying compressive sensing for problems with partially known support,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT), Seoul, Korea, 2009*.
- [30] V. Stankovic, L. Stankovic, and S. Cheng, “Sparse Signal Recovery With Side Information,” 2009.
- [31] R. von Borries, C. Miosso, and C. Potes, “Compressed sensing using prior information,” in *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pp. 121–124, 2007.
- [32] L. Gan, “Block compressed sensing of natural images,” in *15th Int. Conf. on Digital Signal Process.*, pp. 403–406, Jul 2007.
- [33] H. Yap, A. Eftekhari, M. Wakin, and C. Rozell, “The restricted isometry property for block diagonal matrices,” in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pp. 1–6, IEEE, 2011.
- [34] N. Rahnavard, A. Talari, and B. Shahrabi, “Nonuniform compressive sensing,” *49th Annual Allerton Conference on Communication, Control, and Computing*, September 2011.
- [35] C. Luo, F. Wu, J. Sun, and C. W. Chen, “Efficient measurement generation and pervasive sparsity for compressive data gathering,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3728–3738, december 2010.

- [36] M. Crouse, R. Nowak, and R. Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE Trans. Signal Process.*, vol. 46, pp. 886–902, Apr 1998.
- [37] R. Gonzalez and R. Woods, “Digital image processing 3rd edition/e?book images downloads.”
- [38] R. Gonzalez and R. Woods, “Digital image processing—addison wesley publishing company,” 2008.
- [39] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, pp. 4655–4666, Dec 2007.
- [40] “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.
- [41] C. La and M. Do, “Tree-based Algorithms for Compressed Sensing with Sparse-Tree Prior,” March 2009.
- [42] D. Baron, M. Duarte, S. Sarvotham, M. Wakin, and R. Baraniuk, “An information-theoretic approach to distributed compressed sensing,” in *Allerton Conference on Communication, Control, and Computing*, 2005.
- [43] D. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. of the Nat. Academy of Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [44] L. He and L. Carin, “Exploiting structure in wavelet-based bayesian compressive sensing,” *IEEE Trans. Signal Process.*, vol. 57, pp. 3488–3497, Sept 2009.
- [45] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb 2011.

- [46] S. Rangan, P. Schniter, E. Riegler, A. Fletcher, and V. Cevher, “Fixed points of generalized approximate message passing with arbitrary matrices,” in *IEEE Int. Symp. on Inf. Theory Proc. (ISIT), 2013*, pp. 664–668, Jul 2013.
- [47] B. Shahrasbi and N. Rahnavard, “Model-Based Nonuniform Compressive Sampling and Recovery of Natural Images Utilizing a Wavelet-Domain Universal Hidden Markov Model ,” *submitted to IEEE Transactions on Signal Processing*, 2014.
- [48] B. Shahrasbi, A. Talari, and N. Rahnavard, “TC-CSBP: Compressive sensing for time-correlated data based on belief propagation,” *2011 Annual Conference on Information Sciences and Systems (CISS’11)*, March 2011.
- [49] P. Schniter, “Turbo reconstruction of structured sparse signals,” in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1–6, March 2010.
- [50] M. Duarte and Y. Eldar, “Structured compressed sensing: From theory to applications,” *IEEE Trans. Signal Process.*, vol. 59, pp. 4053–4085, Sept 2011.
- [51] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *Signal Processing, IEEE Transactions on*, vol. 56, pp. 2346–2356, June 2008.
- [52] R. Baraniuk, “Compressive sensing,” *Lecture notes in IEEE Signal Processing magazine*, vol. 24, no. 4, pp. 118–120, 2007.
- [53] E. Candès, “Compressive sampling,” in *Proc. of the Int. Congr. of Mathematicians*, vol. 1, Citeseer, 2006.
- [54] E. Candes, M. Rudelson, T. Tao, and R. Vershynin, “Error correction via linear programming,” in *Annual Symposium on Foundations of Computer Science*, vol. 46, p. 295, IEEE Computer Society Press, 2005.

- [55] E. Candes and T. Tao, “The Dantzig selector: statistical estimation when p is much larger than n ,” *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [56] M. Duarte, M. Wakin, and R. Baraniuk, “Fast reconstruction of piecewise smooth signals from incoherent projections,” *SPARS’05*.
- [57] S. Sarvotham, D. Baron, and R. Baraniuk, “Compressed sensing reconstruction via belief propagation,” *preprint*, 2006.
- [58] C. J. Miosso, R. von Borries, M. Argàez, L. Velazquez, C. Quintero, and C. M. Potes, “Compressive sensing reconstruction with prior information by iteratively reweighted least-squares,” *Trans. Sig. Proc.*, vol. 57, no. 6, pp. 2424–2431, 2009.
- [59] N. Vaswani, “LS-CS-residual (LS-CS): Compressive sensing on least squares residual,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4108–4120, 2010.
- [60] N. Vaswani, “Analyzing least squares and kalman filtered compressed sensing,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 3013–3016, 2009.
- [61] N. Vaswani, “Kalman filtered compressed sensing,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2008.
- [62] D. Angelosante, G. B. Giannakis, and E. Grossi, “Compressed sensing of time-varying signals,” in *Digital Signal Processing, 2009 16th International Conference on*, pp. 1–8, July 2009.
- [63] D. Baron, M. Duarte, S. Sarvotham, M. Wakin, and R. Baraniuk, “An information-theoretic approach to distributed compressed sensing,” in *Allerton Conf. Comm., Control, Comput*, Citeseer, 2005.

- [64] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, “Distributed compressed sensing,” 2005.
- [65] S. Madden, “UC-Berkeley Intel Lab Data.” <http://db.csail.mit.edu/labdata/labdata.html>, June 2004.
- [66] F. Kschischang, B. Frey, and H. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Transactions on information theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [67] V. Krishnamurthy and J. Moore, “On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure,” *IEEE Transactions on Signal Processing*, vol. 41, no. 8, pp. 2557–2573, 1993.
- [68] D. Baron and S. Sarvotham, “Compressive sensing via belief propagation software,” December 2008. <http://www.ece.rice.edu/drorb/CSBP/>.
- [69] Y. Oike and A. El Gamal, “CMOS image sensor with per-column $\Sigma\Delta$ ADC and programmable compressed sensing,” *IEEE Journal of Solid-State Circuits*, vol. 48, pp. 318–328, Jan 2013.
- [70] M. Duarte, M. Wakin, and R. Baraniuk, “Wavelet-domain compressive signal reconstruction using a Hidden Markov Tree model,” in *IEEE Int. Conf. on Acoust., Speech and Signal Process., (ICASSP 2008)*, pp. 5137–5140, Mar 2008.
- [71] R. Berinde and P. Indyk, “Sparse recovery using sparse random matrices,” *preprint*, 2008.
- [72] A. Gilbert and P. Indyk, “Sparse recovery using sparse matrices,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, 2010.
- [73] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, “Efficient and robust compressed sensing using optimized expander graphs,” *IEEE Transactions on Information Theory*, vol. 55, pp. 4299–4308, sept. 2009.

- [74] H. Kung and S. Tarsa, "Partitioned compressive sensing with neighbor-weighted decoding," in *MILITARY COMMUNICATIONS CONFERENCE, 2011 - MILCOM 2011*, pp. 149–156, nov. 2011.
- [75] Y. Yu, B. Wang, and L. Zhang, "Saliency-based compressive sampling for image signals," *Signal Processing Letters, IEEE*, vol. 17, pp. 973–976, Nov 2010.
- [76] Y. Shen, W. Hu, R. Rana, and C. T. Chou, "Nonuniform compressive sensing for heterogeneous wireless sensor networks," vol. 13, pp. 2120–2128, June 2013.
- [77] Y. Liu, X. Zhu, L. Zhang, and S. H. Cho, "Expanding window compressed sensing for non-uniform compressible signals," *Sensors*, vol. 12, pp. 13034–13057, Sep 2012.
- [78] N. Rahnavard, B. Vellambi, and F. Fekri, "Rateless codes with unequal error protection property," *IEEE Transactions on Information Theory*, vol. 53, pp. 1521–1532, april 2007.
- [79] M. Khajehnejad, A. Dimakis, W. Xu, and B. Hassibi, "Sparse recovery of nonnegative signals with minimal expansion," *IEEE Transactions on Signal Processing*, vol. 59, pp. 196–208, Jan 2011.
- [80] B. Bah and J. Tanner, "Vanishingly sparse matrices and expander graphs, with application to compressed sensing," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7491–7508, 2012.
- [81] M. S. Pinsker, "On the complexity of a concentrator," in *7th International Teletraffic Conference*, 1973.
- [82] S. Borman and R. L. Stevenson, "Super-resolution from image sequences-a review," in *mws-cas*, p. 374, IEEE, 1998.

- [83] S. Wang, B. Shahrasbi, and N. Rahnavard, “Srl1: Structured reweighted ℓ_1 minimization for compressive sampling of videos,” in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 301–305, Jul. 2013.
- [84] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Advances in neural information processing systems*, pp. 545–552, 2006.
- [85] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21.”
- [86] M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control* (V. Blondel, S. Boyd, and H. Kimura, eds.), Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-Verlag Limited, 2008.
- [87] A. Schulz, L. Velho, and E. A. B. Da Silva, “On the empirical rate-distortion performance of compressive sensing,” in *16th IEEE Int. Conf. on Image Process. (ICIP)*, pp. 3049–3052, Nov 2009.
- [88] P. Indyk and I. Razenshteyn, “On model-based RIP-1 matrices,” in *Automata, Languages, and Programming* (F. Fomin, R. Freivalds, M. Kwiatkowska, and D. Peleg, eds.), vol. 7965 of *Lecture Notes in Computer Science*, pp. 564–575, Springer Berlin Heidelberg, 2013.
- [89] J. Romberg, H. Choi, and R. Baraniuk, “Bayesian wavelet-domain image modeling using hidden markov trees,” in *Int. Conf. on Image Process. (ICIP)*, vol. 1, pp. 158–162, Oct 1999.
- [90] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, “Toeplitz compressed sensing matrices with applications to sparse channel estimation,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 5862–5875, Nov 2010.
- [91] Y. Tsaig and D. Donoho, “Extensions of compressed sensing,” *Signal Process.*, vol. 86, no. 3, pp. 549 – 571, 2006.

- [92] H. Lee, H. Oh, S. Lee, and A. Bovik, “Visually weighted compressive sensing: Measurement and reconstruction,” *IEEE Trans. Image Process.*, vol. 22, pp. 1444–1455, Apr 2013.
- [93] T. Do, L. Gan, N. Nguyen, and T. Tran, “Fast and efficient compressive sensing using structurally random matrices,” *IEEE Trans. Signal Process.*, vol. 60, pp. 139–154, Jan 2012.
- [94] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, “Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images,” *IEEE Trans. Image Process.*, vol. 21, pp. 130–144, Jan 2012.
- [95] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, “Proximal methods for hierarchical sparse coding,” *J. Mach. Learning Research*, vol. 12, pp. 2297–2334, Jul 2011.
- [96] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse problems*, vol. 23, no. 3, p. 969, 2007.
- [97] G.-S. Cheon and B. L. Shader, “Sparse orthogonal matrices and the haar wavelet,” *Discrete Applied Math.*, vol. 101, no. 1/3, pp. 63 – 76, 2000.
- [98] T. Do and B. Mark, “Joint spatial-temporal spectrum sensing for cognitive radio networks,” *IEEE Transactions on Vehicular Technology.*, vol. 59, pp. 3480–3490, Sept 2010.
- [99] Q. Wu, G. Ding, J. Wang, and Y.-D. Yao, “Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: Two-dimensional sensing,” *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 516–526, February 2013.
- [100] S. Manaffam, M. Razeghi-Jahromi, and A. Seyedi, “Stabilizing a random dynamics network with a random communications network,” in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 746–751, Dec 2012.

- [101] C. han Lee and W. Wolf, "Multiple access-inspired cooperative spectrum sensing for cognitive radio," in *IEEE Military Communications Conference, 2007. MILCOM 2007.*, pp. 1–6, Oct 2007.
- [102] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 118–129, Jan 2008.
- [103] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, pp. 1–9, April 2010.
- [104] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, pp. 1902–1916, March 2013.
- [105] Z. Wang, Z. Feng, and P. Zhang, "An iterative hungarian algorithm based coordinated spectrum sensing strategy," *IEEE Communications Letters*, vol. 15, pp. 49–51, Jan 2011.
- [106] J. Oksanen, J. Lundén, and V. Koivunen, "Reinforcement learning based sensing policy optimization for energy efficient cognitive radio networks," *Neurocomputing*, vol. 80, pp. 102–110, Mar 2012.
- [107] A. W. Min and K. G. Shin, "Impact of mobility on spectrum sensing in cognitive radio networks," in *Proceedings of the 2009 ACM Workshop on Cognitive Radio Networks, CoRoNet '09*, (New York, NY, USA), pp. 13–18, ACM, 2009.
- [108] W. Ren, Q. Zhao, and A. Swami, "Temporal traffic dynamics improve the connectivity of ad hoc cognitive radio networks," *IEEE/ACM Trans. Netw.*, vol. 22, pp. 124–136, Feb. 2014.

- [109] G. Caso, H. Soleimani, L. De Nardis, A. Tosti, and M. Di Benedetto, "Sensic: Mobility-aware cluster-based cooperative spectrum sensing for cognitive radio networks," in *Ultra-WideBand (ICUWB), 2014 IEEE International Conference on*, pp. 102–107, Sept 2014.
- [110] B. Shahrabi and N. Rahnavard, "A clustering-based coordinated spectrum sensing in wideband large-scale cognitive radio networks," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 1101–1106, Dec 2013.
- [111] M. Bradonjic and L. Lazos, "Graph-based criteria for spectrum-aware clustering in cognitive radio networks," *Ad Hoc Networks*, vol. 10, pp. 75 – 94, Jan 2012.
- [112] J. Wei and X. Zhang, "Energy-efficient distributed spectrum sensing for wireless cognitive radio networks," in *INFOCOM IEEE Conference on Computer Communications Workshops*, 2010, pp. 1–6, Mar 2010.
- [113] Y. Liu, S. Xie, R. Yu, Y. Zhang, and C. Yuen, "An efficient mac protocol with selective grouping and cooperative sensing in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, pp. 3928–3941, Oct 2013.
- [114] S. Liu, I. Ahmad, Y. Bai, Z. Feng, Q. Zhang, and Y. Zhang, "A novel cooperative sensing based on spatial distance and reliability clustering scheme in cognitive radio system," in *IEEE 78th Vehicular Technology Conference (VTC Fall), 2013*, pp. 1–5, Sept 2013.
- [115] Y. Sun, H. Hu, F. Liu, H. Yi, and X. Wang, "Selection of sensing nodes in cognitive radio system based on correlation of sensing information," in *4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08.*, pp. 1–6, Oct 2008.
- [116] H. Liu and W. Chen, "Cooperative spectrum sensing and weighted-clustering algorithm

- for cognitive radio network,” *IJ Information Engineering and Electronic Business*, vol. 2, pp. 20–27, Mar 2011.
- [117] S. Manaffam and A. Seyedi, “Synchronization probability in large complex networks,” *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 60, pp. 697–701, Oct 2013.
- [118] H. Kim and K. Shin, “Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks,” *IEEE Transactions on Mobile Computing*, pp. 533–545, May 2008.
- [119] E. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, “Cooperative spectrum sensing in cognitive radio networks with weighted decision fusion schemes,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 3838–3847, Dec 2010.
- [120] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, “IEEE 802.22: the first worldwide wireless standard based on cognitive radios,” in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, pp. 328–337, Nov 2005.
- [121] W. Dai, Y. Gai, and B. Krishnamachari, “Online learning for multi-channel opportunistic access over unknown markovian channels,” in *Sensing, Communication, and Networking (SECON), 2014 Eleventh Annual IEEE International Conference on*, pp. 64–71, June 2014.
- [122] R. García-Pelayo, “Distribution of distance in the spheroid,” *Journal of Physics A: Mathematical and General*, vol. 38, no. 16, p. 3475, 2005.
- [123] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

- [124] R. Diestel, “Graph theory, volume 173 of Graduate Texts in Mathematics,” *Springer, Heidelberg*, vol. 91, p. 92, 2005.
- [125] F. Digham, M.-S. Alouini, and M. K. Simon
- [126] A. Abdi, C. Tepedelenlioglu, M. Kaveh, and G. Giannakis, “On the estimation of the k parameter for the rice fading distribution,” *IEEE Communications Letters*, vol. 5, pp. 92–94, Mar 2001.
- [127] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, “IEEE 802.22: The first cognitive radio wireless regional area network standard,” *IEEE Communications Magazine*, vol. 47, pp. 130–138, Jan 2009.
- [128] D. Xue, E. Ekici, and M. C. Vuran, “CORN2: Correlation-based cooperative spectrum sensing in cognitive radio networks,” in *10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012*, May 2012.
- [129] X. Li, S. Hong, Z. Han, and Z. Wu, “Bayesian compressed sensing based dynamic joint spectrum sensing and primary user localization for dynamic spectrum access,” in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, Dec 2011.
- [130] J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, “A primer on spatial modeling and analysis in wireless networks,” *Communications Magazine, IEEE*, vol. 48, pp. 156–163, November 2010.