Compressive Sensing of Multiband Spectrum towards Real-World Wideband Applications

by

Haoran Qi

A thesis submitted to the University of London for the degree of
Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London
United Kingdom

March 2021

Abstract

Spectrum scarcity is a major challenge in wireless communication systems with their rapid evolutions towards more capacity and bandwidth. The fact that the real-world spectrum, as a finite resource, is sparsely utilized in certain bands spurs the proposal of spectrum sharing. In wideband scenarios, accurate real-time spectrum sensing, as an enabler of spectrum sharing, can become inefficient as it naturally requires the sampling rate of the analog-to-digital conversion to exceed the Nyquist rate, which is resource-costly and energy-consuming. Compressive sensing techniques have been applied in wideband spectrum sensing to achieve sub-Nyquist-rate sampling of frequency sparse signals to alleviate such burdens.

A major challenge of compressive spectrum sensing (CSS) is the complexity of the sparse recovery algorithm. Greedy algorithms achieve sparse recovery with low complexity but the required prior knowledge of the signal sparsity. A practical spectrum sparsity estimation scheme is proposed. Furthermore, the dimension of the sparse recovery problem is proposed to be reduced, which further reduces the complexity and achieves signal denoising that promotes recovery fidelity. The robust detection of incumbent radio is also a fundamental problem of CSS. To address the energy detection problem in CSS, the spectrum statistics of the recovered signals are investigated and a practical threshold adaption scheme for energy detection is proposed. Moreover, it is of particular interest to seek the challenges and opportunities to implement real-world CSS for systems with large bandwidth. Initial research on the practical issues towards the real-world realization of wideband CSS system based on the multicoset sampler architecture is presented.

In all, this thesis provides insights into two critical challenges - low-complexity sparse recovery and robust energy detection - in the general CSS context, while also looks into some particular issues towards the real-world CSS implementation based on the

multicoset sampler.

Acknowledgments

Firstly, I would like to give my foremost acknowledgement to China Scholarship Council for providing stipends during my Ph.D. study.

Upon completion of this thesis, I would like to express my sincere thankfulness to those who worked with me in the past years, especially

- Prof. Yue Gao, for his invaluable inspirations, guidance and mentoring throughout the entire time of my Ph.D. without which my Ph.D work would not be possible;
- Prof. Andrea Cavallaro, Dr. Gokop Goteng, Dr. Zhijin Qin, Dr. Xingjian Zhang and Dr. Yuan Ma for their constructive advice and comments on my research work;
- Mr. Peter Anderson for his technical resources and insightful suggestions on hardware development, and laboratory managers, Mr. Kok Ho Huen and Dr. Max Munoz Torrico for their cordial assistance on my practical work;
- supportive teams at QMUL, especially Ms. Melissa Yeo and the EECS staff, for their enthusiastic support for the student community, especially during the disruption of COVID-19;
- my dear colleagues, Dr. Shaker Alkaraki, Dr. Qianyun Zhang, Mr. Wei Su, Dr. Andrea Toma, Dr. Kang Liu, Dr. Zixiang Ma, Dr. Mingtuan Lin, Ms. Xiaolan Liu, Ms. Jiadong Yu, Mr. Zihang Song and all, who demonstrated the spirit of collaboration.

At last, with great love, I would like to give my special gratitude to my family who constantly devote their unconditional love and support throughout my life, and my dear friends who share my up and downs all the way.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3GPP	third generation project program
5G	the fifth generation
ADC	analog-to-digital converter
ADMM	alternating direction method of multipliers
AIC	analog-to-information converter
AWGN	additive wideband Gaussian noise
BIC	Bayesian information criterion
BICe	enhanced Bayesian information criterion
BP	basis pursuit
BPF	bandpass filter
CBRS	citizens broadband radio service
CCPD	complementary cumulative probability density
CFAR	constant false alarm rate
CFD	cyclostationary feature extraction
CLT	central limit theorem
CoSaMP	compressive sampling matching pursuit
CR	cognitive radio
CS	compressive sensing
CSS	compressive spectrum sensing
DFT	discrete Fourier transform
DMA	direct memory access
DoF	degree of freedom

DSA	dynamic spectrum access
DTFT	discrete-time Fourier transform
DTTV	digital terrestrial television
ED	energy detection
EIRP	equivalent isometry radiation power
EM	expectation-maximazation
ESR	error-to-signal ratio
FCC	federal commission of communications
FFT	fast Fourier transform
FIFO	first-in-first-out
FOCUSS	focal underdetermined system solver
FPC	fixed-point continuation
FPGA	field programmable gate array
FR	frequency range
GSM	global systems for mobile communications
GUI	graphical user interface
HD	hard detection
HTP	hard thresholding pursuit
IDFT	inverse discrete Fourier transform
JB-HTP	joint-block HTP
IEEE	institution of electrical and electronic engineers
IF	intermediate frequency
IoT	Internet-of-Things
IRLS	iterative regularized least-squares
ISM	industrial, scientific and medical
JB-HTP	joint block hard thresholding pursuit
KKT	Karush–Kuhn–Tucker
KLD	Kullbeck-Liebler divergence
LAA	licence-assisted access

LARS	least angle regression
LASSO	least absolute shrinkage and selection operator
LBT	listen-before-talk
LPWAN	low-power wide-area network
LTE	long-tern evolution
LTE-U	long-tern evolution unlicensed
M2M	machine-to-machine
MDL	minimum description length
ML	maximum likelihood
MM	mixture model
MMV	multiple measurement vector
mmWave	millimeter wave
MWC	modulated wideband converter
NI	National Instruments
NMSE	normalised mean-squared error
NP	Neyman-Pearson
NR	new radio
NR-U	new radio unlicensed
OfCom	Office of Communications
OFDM	orthogonal frequency-division multiplexing
OMP	orthogonal matching pursuit
PAN	personal area network
PCIe	peripheral component interconnect express
PDF	probability density function
PLL	phase-locked loop
PSD	power spectrum density
PU	primary user
RDPI	random demodulator pre-integrator
RF	radio frequency

RIP	restricted isometry property
ROC	receiver operating characteristics
SA-SOMP	subspace augmented simultaneous orthogonal matching pursuit
SAS	spectrum access service
SBL	sparse Bayesian learning
SDR	software-defined radio
SMV	single measurement vector
SNR	signal-to-noise ratio
SOMP	simultaneous orthogonal matching pursuit
SP	subspace pursuit
SPG	spectral projected gradient
StOMP	stagewise orthogonal matching pursuit
SU	secondary user
TVWS	TV whitespace
UHF	ultra high frequency
WRAN	wireless regional area network
WiGig	Wireless Gigabit

About Mathematical Symbols

Conventions

Scalars are denoted normal-font symbols, e.g. B for bandwidth;

Vectors are denoted by lowercase **bold**-font symbols, e.g. $\boldsymbol{b}$ for vector of additive noise;

Matrices are denoted by UPPERCASE **bold**-font symbols, e.g. $\boldsymbol{A}$ for sensing matrix.

List of Key Mathematical Symbols

Definitions of symbols are available in the context. This is a list of key mathematical symbols in all chapters for readers' reference. For conciseness, some less significant symbols only appearing in the local context of its definition are not listed here.

$\boldsymbol{A}$	sensing matrix
$\boldsymbol{A}_\tau$	desired-offset-related component of sensing matrix in multicoset sampler model
$\boldsymbol{A}_\delta$	undesired-skew-related component of sensing matrix in multicoset sampler model
α	a parameter of the energy statistical model in (4.10)
B	baseband bandwidth
$\boldsymbol{b}$	additive noise in single-vector form
$\boldsymbol{B}$	additive noise in multiple-vector form
$\boldsymbol{B}_v$	additive noise in dimension-reduced multiple-vector form
β	a parameter of the energy statistical model in (4.10)

C	number of channels
c_p	delay in number of Nyquist sampling interval for the pth branch of the multicoset sampler
$\boldsymbol{\gamma}$	hyperparameters of sparse Bayesian Learning algorithm
$C^{'}$	number of vacant channels determined by the threshold adaption scheme in Chapter 4
$\mathcal{C}$	set of vacant channels determined by the threshold adaption scheme in Chapter 4
$\delta_.$	restricted isometry constant
$\delta(\cdot)$	sparsity of a vector or row sparsity of a matrix
δ_p	undesired skews for the p the channel of multicoset sampler
$\tilde{\boldsymbol{\Delta}}[k]$	the diagnol matrix in practical multicoset sampler model
$\tilde{\boldsymbol{\Delta}}_\tau[k]$	desired-offset-related of the diagnol matrix in practical multicoset sampler model
$\tilde{\boldsymbol{\Delta}}_\delta[k]$	undesired-skew-related of the diagnol matrix in practical multicoset sampler model
ϵ	constant in the l_1 minimization problem (2.10)
f_{NYQ}	Nyquist rate
f_{LAU}	Landau's rate
f_s	sampling frequency
F_s	OFDM symbol rate
$f(\cdot)$	probability density function
$\boldsymbol{\mathcal{F}}$	Fourier basis, i.e. DFT matrix
$\rho_1, \rho_2, \psi_1, \psi_2$	parameter notations defined in (5.68) for multicoset sampler model
$\mathcal{H}_0$	the null hypothesis, i.e. channel vacant
$\mathcal{H}_1$	the alternative hypothesis , i.e. channel occupied
$\boldsymbol{I}(\cdot,\cdot)$	Fisher information matrix
$(\cdot)^{<j>}$	parameters in the jth iteration of the algorithms in Chapter 4
k^o	the estimated sparsity by the BICe criteria in Chapter 3
k^c	the compensated sparsity estimation in Chapter 3
$\mathcal{K}$	the set of occupied channels
λ	constant in the l_1 minimization problem (2.12)
$\lambda_.$	eigenvalues of autocorrelated measurements in noiseless case
$\hat{\lambda}_.$	eigenvalues of the autocorrelated measurements in general case
$\hat{\lambda}_.^{'}$	eigenvalues of the autocorrelated measurements for dimension reduction
l	actual number of occupied channels
l^o	estimated number of occupied channels in Chapter 3
$\boldsymbol{\Lambda}_s$	a diagonal matrix containing descending-ordered signal-subspace eigenvalues
$\boldsymbol{\Lambda}_n$	a diagonal matrix containing descending-ordered noise-subspace eigenvalues
L	number of observations of MMV measurements
M	number of compressed measurements in a measurement vector

N	number of samples in the (Nyquist-rate) orignal signal vector
$\boldsymbol{n}_f$	frequency-domain additive noise for non-compressive spectrum sensing model
$\boldsymbol{n}_t$	time-domain additive noise for non-compressive spectrum sensing model
P	number of parallel branches for multicoset sampler or modulated wideband converter
$\pi_{\mathcal{H}_0}$	prior probability of a sample belonging to the null hypothesis in Chapter 4
$\pi_{\mathcal{H}_1}$	prior probability of a sample belonging to the alternative hypothesis in Chapter 4
$p_c^{(c)}$	average of PSD values within the cth channel
p_c	average of PSD values within one channel
$p_c[t]$	the tth sample of p_c
p_o	the estimated row sparsity of the sliced spectrum $\hat{\boldsymbol{X}}$ in multicoset sampler
σ^2	noise variance
$Q(\cdot,\cdot)$	surrogate function in EM algorithm
$\boldsymbol{R}_y$	autocorrelation of measurement vectors
$\boldsymbol{R}_{s_f}$	autocorrelation of orignal sparse signals
$\hat{\boldsymbol{R}}_y$	estimated autocorrelation of measurement vectors
$\hat{\boldsymbol{R}}_{s_f}$	estimated autocorrelation of orignal sparse signals
$\hat{\boldsymbol{R}}_y^{'}$	estimated autocorrelation of measurement vectors for dimension reduction
$\hat{\boldsymbol{R}}_{s_f}^{'}$	estimated autocorrelation of original sparse signals for dimension reduction
$\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}$	estimated autocorrelation of measurements in MMV model
$\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{X}}}}$	estimated autocorrelation of orignal row-sparse signals in MMV model
R	a parameter of the energy statistical model in (4.10)
$s(t)$	continuous-time signal at baseband
$s[n]$	discrete-time signal at baseband
$\boldsymbol{s}$	sparse signal in general
$\boldsymbol{s}^o$	recovered sparse signal by optimization
$\boldsymbol{s}_t$	time-domain presentation of the original signal
$\boldsymbol{s}_f$	original frequency-domain sparse signal
$\boldsymbol{s}_f^r$	recovered frequency-domain sparse signal
$s_{c_p}[n]$	discrete-time signal sampled by the pth branch of multicoset sampler
$\tilde{s_{c_p}}[n]$	finite-length signal sampled by the pth branch of practical multicoset sampler
$\boldsymbol{S}$	sparse signal in multiple measurement model in general
$\boldsymbol{S}_f$	original frequency-domain sparse signal in MMV model

$\boldsymbol{S}_f^{'}$	original frequency-domain sparse signal in MMV model
$\boldsymbol{S}_f^{'\,r}$	recovered frequency-domain sparse signal in MMV model
$\boldsymbol{S}_v$	original frequency-domain sparse signal in dimension-reduced MMV model
$\boldsymbol{S}_v^r$	recovered frequency-domain sparse signal in dimension-reduced MMV model
$\text{supp}(\cdot)$	the support of a vector (set of indices of non-zero elements) or matrix (set of indices of non-zero rows)
T_s	sampling interval
$\boldsymbol{t}$	Lagrange multiplier
$\boldsymbol{t}^o$	the optimum of Lagrange multiplier
T	number of single-measurement-vector observations
$T^{'}$	number of single-measurement-vector observations for dimension reduction
$\boldsymbol{\theta}$	set of all parameters in the energy statistical model in Chapter 4
$\boldsymbol{\theta}^o$	set of all optimal parameters in the energy statistical model in Chapter 4
τ_p	desired offsets for the p the channel of multicoset sampler
$\mu(\cdot)$	mutual cohenrence of a matrix
$\boldsymbol{v}_.$	eigenvectors of autocorrelated measurements in noiseless case
$\hat{\boldsymbol{v}}_.$	eigenvectors of the autocorrelated measurements in general case
$\hat{\boldsymbol{v}}_.^{'}$	eigenvectors of the autocorrelated measurements for dimension reduction
$\boldsymbol{V}_s$	a matrix consisting of eigenvectors spanning the signal subspace
$\boldsymbol{V}_n$	a matrix consisting of eigenvectors spanning the noise subspace
$\boldsymbol{\Phi}$	sampling matrix
W	number of spectrum bins corresponding to the main lobe of Dolph-Chebyshev window
$X(f)$	Fourier transform of the continuous-time baseband signal $s(t)$
$X_{c_p}(f)$	discrete-time Fourier transform of the signal $s_{c_p}[n]$
$\tilde{X}(f)$	Fourier transform of the windowed continuous-time signal $\tilde{s}(t)$
$\boldsymbol{x}$	original signal in general
$\boldsymbol{x}_f$	frequency-domain sensed signal for non-compressive spectrum sensing model
$\boldsymbol{x}_t$	time-domain sensed signal for non-compressive spectrum sensing model
$\boldsymbol{x}(f)$	continuous spectrum sliced into channel in multicoset sampler model
$\tilde{\boldsymbol{x}}[k]$	discrete spectrum sliced into channel in practical multicoset sampler model
$\tilde{\boldsymbol{X}}$	sliced spectrum in matrix form in practical multicoset sampler model
$\hat{\boldsymbol{X}}$	sliced spectrum in MMV model without significant spectrum leakage components
ξ	constant in the l_1 minimization problem (2.11)
$\boldsymbol{\Psi}$	sparsifying matrix
$\boldsymbol{y}$	measurement vector
$\boldsymbol{y}(f)$	measurement vector in multicoset sampler model
$\tilde{\boldsymbol{y}}[k]$	measurement vector in practical multicoset sampler model
$\boldsymbol{Y}$	measurements in MMV model in general case
$\boldsymbol{Y}_s$	measurements in MMV model in noiseless case

$\tilde{\boldsymbol{Y}}$	matrix consisting multiple measurement vectors $\tilde{\boldsymbol{y}}[k]$ for all k's
$\hat{\boldsymbol{Y}}$	measurements in MMV model without significant spectrum leakage components
$\omega(t)$	window function in general
$\omega_{\mathrm{dch}}[n]$	Dolph-Chebyshev window function in discrete time form
$\omega_{\mathrm{rect}}[n]$	rectangular window function in discrete time form
Ω	set of detected block entries by greedy pursuit

Chapter 1

Introduction

The rapid evolution of wireless communications has demanded an increasingly large data rate and service coverage, and the spectrum scarcity has been increasingly appearing to be a major challenge of wireless communication applications [1]. In recent years, there has been particular interest in the techniques of spectrum sharing leading to the more efficient utilization of the frequency resources which have become major commercial assets of wireless service operators. Rapid technical advances towards the deployment of dynamic spectrum access (DSA) have been seen through operation models such as geolocation database and cognitive radio techniques [2, 3].

The initial trial steps of DSA systems are led by regulation bodies to tackle the crucial threat of spectrum scarcity. A real-world test case for DSA in the UK is TV white-space (TVWS) where the allocated spectrum of $470 - 790$ MHz has been largely vacated by the massive introduction of digital terrestrial TV (DTTV) [4–6]. The TVWS lying in the lower end of the ultra-high frequency (UHF) band and featuring wide bandwidth and sparse spectrum usage by primary TV broadcast is particularly suitable for the implementation of DSA techniques [7], especially for secondary low-power machine-to-machine (M2M) type communications. The Federal Communications Commission (FCC) of the US has also initiated DSA in TVWS bands [8] as well as in the underutilized 3.5

GHz band previously used exclusively by US federal services [9].

To enable the reliable awareness of the ambient spectrum usage required by DSA, cognitive radio techniques are attractive options. Spectrum sensing is the essential and proactive way in cognitive radio to acquire the ambient spectrum availability information, which enables dynamic access demanded by unlicensed secondary users (SUs) to the spectrum, while guaranteeing the quality of service of licensed primary users (PUs) [10, 11]. The crucial requirement of DSA to reliably obtain all dynamics of the spectrum activities, and to that end, real-time spectrum sensing techniques are highly desired. Nyquist-rate sampling is an essential underlying requirement in conventional real-time spectrum sensing, presenting high system complexity and cost, especially when the interested band is wide. Compressive sensing (CS) techniques have been recently proposed as promising candidates in wideband spectrum sensing to achieve sub-Nyquist-rate sampling, which effectively alleviates the hardware and power constraints of the conventional Nyquist-rate processing [12–16].

With the recent commercialization of the fifth-generation (5G) communication system, the spectrum designation not only refers to the sub-6 GHz band and some extensions to 7125 MHz, known as frequency range 1 (FR1) [17], but also new spectrum allocations FR2 in the millimeter-wave (mmWave) frequencies between 24.25 GHz and 52.6 GHz [18]. Although the mmWave spectrum resources for civil cellular services are new and large, in consideration of the coexistence of multiple operators and various types of applications, high spectrum utility and effective spectrum access management are necessary, and spectrum sharing is considered as a promising model. Recent proposals are present that cellular and machine-to-machine-type services at mmWave bands, with either unlicensed and shared spectrum in definition ($37.0-37.6$ GHz and $64-71$ GHz) [19, 20], or with multiple operators potentially servicing overlapping small areas and co-existing with other crucial wireless services such as Ka-band satellite communication links ($27.5-29.5$ GHz, $37-39$ GHz, etc.) [20], could employ spectrum sensing techniques to realize the DSA model or to ensure the quality of service [21]. Real-time spectrum sensing for such

mmWave DSA applications, where the bandwidth can span multiple gigahertz, poses critical challenges on the sampling rate in the conventional context of super-Nyquist rate acquisition. Sub-Nyquist signal processing based on compressive sensing techniques has hence emerged to attention and could be promising candidates for wideband spectrum sensing, which forms the fundamental interest of this thesis.

Implementing compressive spectrum sensing (CSS) has seen various practical and technical challenges. As the core procedure of CSS, the robustness and computational complexity of the sparse recovery process are the major concerns as the opportunistic access of the spectrum requires rapid and reliable knowledge of the spectrum occupancy. While greedy algorithms are preferred in light of achieving relatively low computational complexity, they generally require prior knowledge of the spectrum sparsity as the input, the acquisition of which poses another necessary yet intriguing problem. In this thesis, a practical low-complexity CSS scheme featuring some greedy algorithms is proposed, along with the solution of the sparsity estimation and insights of further techniques to reduce the recovery complexity in Chapter 3. Moreover, the robust detection of incumbent radio activities is also crucial in determining the spectrum availability. In this thesis, the energy detection problem in CSS is investigated and a practical threshold adaption scheme is proposed to ensure the robust performance of spectrum availability detection in Chapter 4.

To apply CSS techniques to systems with larger bandwidth for mmWave applications, where the instantaneous bandwidth reaches multiple hundred megahertz or even several gigahertz, considerations have to be given to the system design as the corresponding Nyquist rate is considerably higher. Some practical limitations of the implementation of CSS, though being less of a problem for relatively narrow bandwidth systems, e.g. the timing precision and spectrum leakage, have been brought to greater attention through the attempts to implement CSS towards mmWave frequencies. This thesis provides practical modeling of the multicoset sampler, then initially looks into several topics of real-world issues in Chapter 5.

1.1 Motivations and Contributions

The summary of the motivations and contributions of this Ph.D. research is given as follows.

Greedy algorithms are of particular interest due to their low complexity but the requirement of prior knowledge of the signal sparsity has been the limitation of their practical implementations. A practical spectrum sparsity estimation scheme and subspace-aided dimension reduction approach for greedy algorithms have been presented to enable low-complexity sparse recovery and improved detection performance of CSS.

The robust detection of incumbent radio is a fundamental problem of CSS toward its application of DSA. It is discovered that the statistical model for Neyman-Pearson (NP) energy detection is different in CSS from conventional Nyquist-rate spectrum sensing. The channel energy statistical model of the spectrum has been investigated in the context of NP energy detection in CSS, followed by a practical threshold adaption scheme proposed to realize the robust detection performance.

It is of particular interest to seek the challenges and opportunities to implement real-world CSS for mmWave systems with larger bandwidth which imposes larger processing complexity and more stringent timing requirements. A few practical issues towards the real-world realization of CSS system with multi-gigahertz bandwidth based on the multi-coset sampler architecture is presented, along with implementation trials on a software-defined-radio (SDR) platform.

1.2 Outline of the Thesis

The remaining chapter of the thesis is organized as the following, with an overview of each chapter's contents.

1.2.1 Background

Chapter 2 covers an introduction to the literature review and background relating to the work in this thesis. Reviews of cognitive radio, dynamic spectrum sharing, and wideband spectrum sensing techniques are presented. Moreover, background of sampling theories, compressive sensing theories and system architectures of CSS is introduced.

1.2.2 Low-Complexity Compressive Spectrum Sensing with Sparsity Estimation

The complexity of the CS recovery algorithm and the detection performance against noise are two of the main challenges of the implementation of CSS. Greedy algorithms have been of particular interest in CSS due to low complexity. However, they require prior knowledge of signal sparsity as input. Existing sparsity estimation schemes require multiple exhaustive pilots of recovery implementations. A novel spectrum sparsity estimation scheme is firstly proposed which directly estimates sub-Nyquist measurements. With such sparsity input, the computational effort of greedy pursuit algorithms can be saved and recovery performance improved, compared to an excessively large input sparsity. Besides, the spectrum sparsity estimates also enable hard detection of channel occupancy where threshold adaption for energy detection is avoided. Moreover, with the detected dimension of signal subspace, it is proposed to implement a multiple-measurement-vector (MMV) model of CSS whose dimension can be reduced, and meanwhile, a large portion of the noise is removed. The proposed MMV model with noise and dimension reduction further improves the detection performance and also keeps the complexity low. Finally, the hard thresholding pursuit (HTP) algorithm is generalized to recover joint-block-sparse MMV signals. In simulations, the detection performance and complexity of the proposed CSS scheme show striking superiority against multiple benchmarking schemes. The work in this chapter has been published as ② and ⑩ in the Publication List.

1.2.3 Channel Energy Modeling and Threshold Adaption in Compressive Spectrum Sensing

Most existing literature discusses Neyman-Pearson channel energy detection and threshold adaption schemes to achieve optimal performance of detection in conventional non-compressive spectrum sensing scenarios. However, in the CSS, it is found that the channel energy statistics and optimal threshold not only depend on noise energy but also compression ratio, the sparsity of spectrum, and the nature of recovery algorithms. The corresponding optimal threshold in energy detection is different from that in the conventional non-compressive spectrum sensing as a result of the discrepancy between the recovered signal and the original sparse signal. To investigate the channel energy statistics of the recovered spectrum, the statistical model of channel energy for CSS is postulated and propose a learning algorithm based on Mixture Model and Expectation-Maximization techniques. Additionally, having verified the validity of the postulated model, a practical threshold adaption scheme for CSS is proposed, aiming to keep constant false alarm rates in channel energy detection. In simulations, it is shown that the postulated channel energy statistical models with parameters learned by the proposed learning algorithm fit well with empirical distributions under circumstances of various channel models and recovery algorithms. Moreover, it is presented that the proposed threshold adaption scheme keeps the false alarm rate near the predefined constant, which in turn validates the postulated model. The work in this chapter has been published as ③ and ⑪ in the Publication List.

1.2.4 Towards Practical Multicoset Sampler for Compressive Spectrum Sensing at mmWave

Multicoset sampler has been widely discussed as a concept in the literature to acquire compressed samples of wideband signals. However, few discussions are seen towards the hardware realization of this architecture. Technical challenges include but are not limited to the precise sampling timing control and clock distribution at Nyquist-rate resolution, and the synchronization and parallel data acquisition from a large-scale analog-to-digital

converter (ADC) array. While the latter one has become less significant thanks to the recent technical advances of JESD204B interface standard for ADC arrays, the timing control and distribution is still a crucial issue.

Firstly, the practical multicoset sampler model is established with considerations on finite length of observation period and discrete samples on spectrum leakage and sparsity estimation. Then, some application-inspired issues of practical multicoset sampler are investigated. Spectrum leakage as a result of the finite observation period affects the sparsity properties of the sensed spectrum, and a specific windowing scheme is proposed to mitigate the spectrum leakage impact on the sparse signal model. Moreover, sparsity estimation is also proposed to be applied on the multicoset model to enable low-complexity greedy algorithms for recovery. With the sparsity estimate, it is presented that the sparse recovery model of multicoset sampler can be further simplified by dimension reduction via eigendecomposition methods. This part of work has been published as ⑨ in the Publication List.

To validate the real-time processing performance required by the multicoset sampler, a full receiver and signal processing testbed of the multicoset-sampler-based CSS system has been developed based on the National Instrument mmWave SDR platform. The processing routine is implemented on the combination of field-programmable gate array (FPGA) target and host processor of the National Instruments mmWave soft-defined-radio platform with an instantaneous bandwidth of 3GHz and the integration of full receiver architecture at 28.5GHz. This part of the work has been published as ⑬ in the Publication List.

To apply practical multicoset sampler towards mmWave bands with even larger bandwidth of multi-gigahertz, the timing offsets of sampling timing should be controlled at finer resolutions and becomes more prone to undesired skews. The impact of timing skews of the multicoset sampler is analyzed, which is unavoidable and has to be taken into account in real-world implementations. The error in compressed measurements caused by timing skews is quantified for the cases of unknown and known skews respec-

tively, which can be indicative of the sparse recovery and system-level performances in the design of practical multicoset-sampler-based wideband spectrum sensing systems. This part of the work has been published as ⑧ in the Publication List.

1.2.5 Conclusions and Future Work

This chapter presents the conclusion and visions for future work.

1.3 Publication List

Patent

① Y. Gao and **H. Qi** (inventors), "Signal Processing Device", UK Patent Application No. 1912010.4, (application filed in August 2019).

Journal Paper

② **H. Qi**, X. Zhang and Y. Gao, "Low-Complexity Subspace-Aided Compressive Spectrum Sensing over Wideband Whitespace", in IEEE Transactions on Vehicular Technology, vol. 68, no. 12, pp. 11762-11777, Dec. 2019.

③ **H. Qi**, X. Zhang and Y. Gao,"Channel Energy Statistics Learning in Compressive Spectrum Sensing," in IEEE Transactions on Wireless Communications, vol. 17, no. 12, pp. 7910-7921, Dec. 2018.

④ X. Zhang, Y. Ma, **H. Qi**, Y. Gao, et. al.,"Distributed Compressive Sensing Augmented Wideband Spectrum Sharing for Cognitive IoT," in IEEE Internet of Things Journal, vol. 5, no. 4, pp. 3234-3245, Aug. 2018.

⑤ X. Zhang, Y. Ma, **H. Qi** and Y. Gao,"Low-Complexity Compressive Spectrum Sensing for Large-Scale Real-Time Processing," in IEEE Wireless Communications Letters, vol. 7, no. 4, pp. 674-677, Aug. 2018.

⑥ X. Liu, J. Yu, **H. Qi**, J. Yang, W. Rong, X. Zhang and Y. Gao, "Learning to

Predict the Mobility of Users in Mobile mmWave Networks," in IEEE Wireless Communications, vol. 27, no. 1, pp. 124-131, Feb. 2020.

(7) A. Toma, A. Krayani, M. Farrukh, **H. Qi**, L. Marcenaro, Y. Gao and C. Regazzoni, "AI-Based Abnormality Detection at the PHY-Layer of Cognitive Radio by Learning Generative Models," in IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 1, pp. 21-34, Mar. 2020.

Conference Paper

(8) **H. Qi** and Y. Gao, "On Timing Skews of Multicoset Samplers in Compressive Spectrum Sensing for Millimeter-Wave", IEEE Global Communication Conference (GLOBECOM), Dec. 2019.

(9) **H. Qi**, J. Yu, A. Cavallaro and Y. Gao, "Robust Compressive Sensing of Multiband Spectrum with Partial and Incorrect Priors," IEEE International Conference on Communications (ICC), May 2019.

(10) **H. Qi**, X. Zhang and Y. Gao, "Subspace-Aided Low-Complexity Blind Compressive Spectrum Sensing over TV Whitespace", IEEE Global Communication Conference (GLOBECOM), Dec. 2018.

(11) **H. Qi**, X. Zhang and Y. Gao, "Channel Energy Statistics Modeling and Threshold Adaption in Compressive Spectrum Sensing", IEEE/CIC International Conference on Communications in China (ICCC), Aug. 2018.

(12) **H. Qi** and Y. Gao, "Two-Dimensional Compressive Spectrum Sensing in Collaborative Cognitive Radio Networks," IEEE Global Communication Conference (GLOBECOM), Dec. 2017.

(13) Z. Song, **H. Qi** and Y. Gao, "Real-time Multi-Gigahertz Sub-Nyquist Spectrum Sensing System for mmWave", ACM Workshop on Millimeter-Wave Networks and Sensing Systems (mmNets), Oct. 2019.

(14) J. Yu, X. Liu, **H. Qi**, W. Zhang and Y. Gao, "Spatial Channel Covariance Estimation for Hybrid mmWave Multi-User MIMO Systems," IEEE Global Communication Conference (GLOBECOM), Dec. 2019.

Chapter 2

Literature Review and Background

The chapter starts with the literature review of cognitive radio, spectrum, and the state-of-the-art of conventional and sub-Nyquist wideband spectrum sensing techniques. Next, with the focus on sub-Nyquist wideband spectrum sensing, the basics of sampling theories and compressive sensing theories are covered, followed by the introductions of compressive sensing algorithms and sub-Nyquist signal acquisition architectures.

2.1 Cognitive Radio and Spectrum Sharing

A cognitive radio (CR), by definition, is a radio that can be programmed and configured dynamically to use the appropriate wireless channels in its vicinity to avoid user interference and congestion. Such a radio can distinguish available channels in the wireless spectrum, then accordingly changes its transmission or reception parameters to allow wireless communications to have more throughput in a given spectrum band at one location. The primary idea of the 'learning machines' for adaptive radio communications was brought up by Shannon. The concept of CR was firstly introduced by Mitola, who proposed a goal towards which the personal wireless devices should evolve: a fully recon-

figurable wireless transceiver that automatically adapts its communication parameters to network and user demands [22, 23]. Recent advances in CR technologies have brought up the concept of dynamic spectrum access (DSA), which allows secondary users (SUs) to opportunistically access the spectrum that is sporadically used by primary users (PUs) in both frequency and time [11, 24]. Such SUs often operate on an unlicensed base and therefore it is required that the access made by SUs to the spectrum should not cause significant interference to the licensed PUs whose quality of service is prioritized. For that purpose, a SU in the sense of cognitive radio should have the ability to obtain the knowledge of the holes, i.e. whitespace in the spectrum at certain times.

Spectrum resources are conventionally allocated by governmental regulators in a rather fixed fashion and are not optimized for CRs. Commercial operations of wireless communication services, in the aspect of spectrum resources usage, can be largely categorized into two: some are based on the monopoly of certain frequency bands at a certain area, for example, radio and TV broadcast and wireless cellular network services such as GSM and LTE; and some others are based on frequency bands allocated for unlicensed use in the industrial scientific and medical (ISM) bands, for example, wireless local area network (WLAN), personal area networks (PAN), and low-power wide-area network (LPWAN) for Internet-of-Things (IoT) applications. Due to the ever-increasing demand for faster data rate and service coverage, the finite resources of the spectrum have become scarce assets and the problem of spectrum scarcity has become critical over the past few years. Nevertheless, it has been reported [2, 4, 10] in several studies that the seemingly overloaded spectrum for some wireless services that are not designed for CR is actually significantly underutilized, in the sense that only a small number of narrowband transmissions occupy across a wideband spectrum. On the other hand, fixed and monopolized spectrum allocation prevents rarely used frequencies, for specific services such as terrestrial TV and paging, from being utilized, even when any SU would not cause significant interference to the primary service. The actual existence of such frequency whitespace enables the potential application model of DSA, where the total

throughput of the wideband spectrum can be significantly improved [24, 25]. Therefore, CR along with the paradigm of spectrum sharing has been considered as one of the promising solutions to tackle the spectrum scarcity in future wireless networks towards 5G and beyond [10].

Facing the urgent need for accessing the shared spectrum to maximize the spectrum utilization, regulation bodies and governments have taken steps to initiate the spectrum sharing model in multiple appropriate bands. One of the first frequency bands that have come to attention is the digital terrestrial TV (DTTV) bands, for the following two reasons. One is that the allocated band is as wide as several hundreds of megahertz which has been largely vacated by the transition from analog to digital TV signals, creating a large portion of whitespace.[4, 26]. Moreover, the frequencies lie in the range of the ultra-high-frequency (UHF) band which has desirable penetration capabilities in terms of the propagation characteristics, especially suitable for long-range, wide-area, and low-power wireless communications such as the rapidly-developing IoT networks [4, 26]. The Federal Communications Commission (FCC) of the United States has enabled the TV whitespace (TVWS) between $54-698$ MHz since 2010 and the Office of Communications (OfCom) of the United Kingdom has permitted $470-790$ MHz since 2015 [5, 6]. Besides regulatory policies, the standardization body IEEE has released two standards dedicated for TVWS applications, namely 802.22 [27] and 802.11af (also known as Super Wi-Fi and White-Fi) [28] for application scenarios of wireless regional area networks (WRAN) with service radius up to 100 km and Wi-Fi services, respectively. Furthermore, in 2015, the FCC adopted the spectrum sharing model for the band of $3550-3700$ MHz, established for the Citizens Broadband Radio Service (CBRS)[29]. Instead of the primitive two-tier model consisting of PUs and SUs, this application of spectrum sharing employs a three-tiered access and authorization framework to accommodate the shared licensed and unlicensed use of the spectrum [9].

Commercial operators also have proposed initiatives to employ a shared spectrum operation model in cellular network services. One of the most attractive proposals is

called LTE-Unlicensed (LTE-U), where the LTE base stations are allowed to operate in or offload data transmissions to the unlicensed bands while still using the unmodified LTE physical and medium access control layer standards [30]. One major proposal is for the ISM band of 5 GHz, where the LTE-U coexists with Wi-Fi services of the IEEE 802.11a/n/ac/ax physical layer standard [31]. LTE-U is intended to facilitate cellular carriers to boost data speeds within small cells and improve the level of network convergence. The 3GPP standardization on LTE-U, namely Licensed Assisted Access (LAA), mandates the LTE-U base station to detect free channels of the spectrum defined in 802.11ac and employ a 'listen-before-talk' (LBT) contention protocol [31]. However, there has been considerable debate over the coexistence of LTE-U and Wi-Fi [32], especially in the United States where LBT is not mandated, that the deployment of LTE-U would severely degrade the performance of Wi-Fi services, and the performance advantage of LTE-U without LAA is limited [30]. With the evolution of cellular network services towards the fifth-generation (5G), the 5G new radio (NR) unlicensed (NR-U) exploiting the concept of LTE-U has been a study item since the standardization of NR Rel-16 [33] to expand the spectrum utility and applicability of 5G NR. It is proposed in the Rel-16 study that standalone NR-U for frequency range (FR) 1 (FR1) is designed to coexist with Wi-Fi and LTE-U LAA employing LBT in the 5 GHz band. Moreover, the anchored NR-U (also noted as carrier aggregation and dual connectivity NR-U) could use licensed spectrum NR/LTE or shared spectrum services such as the 3.5 GHz CBRS as an anchor while offloading to the 5 GHz unlicensed spectrum using LAA method [33]. Work item including LTE-U for millimeter-wave (mmWave) FR2 bands is expected to present in NR Rel-17 [34], and technical discussions have seen to address the distinctive spectrum sharing paradigm due to the highly-directional beam-based transmissions and spatial spectrum reuse in the physical layer, along with the possible coexistence with the unlicensed spectrum service Wireless Gigabit (WiGig) in 60 GHz, i.e. IEEE 802.11ad/ay [35, 36]. Dynamic spectrum sharing has been made available to enable the coexistence of LTE and 5G NR by sharing the carriers in the existing LTE frequencies and such standardization is only limited to FR1. Although absent in current 3GPP

standardization, proposals of shared spectrum models have also been actively considered and evaluated with the opening of mmWave bands to maximize the spectrum utilities by regulation bodies[19, 21]. The FCC has set rules for the licensed and shared use of $37 - 37.6$ GHz for DSA among different commercial users and federal users [20]. Moreover, a total of 14 GHz bandwidth spectrum has also been proposed to be available by the FCC for unlicensed access [37, 38]. The rules also require schemes to be promoted to protect coexisting and spectrum-sharing operations of satellite communication links and other federal services [38]. In particular, the highly directional fixed-satellite service transmissions could also cause the 5G mmWave service coverage outage, and it is critical to identify these satellite services and mitigate the interference not only to the satellite service but also to ensure 5G service quality possibly by null beamforming of the 5G base station and user equipment, whose feasibility is studied in [39, 40].

One of the most crucial capabilities of CR, in the context of shared spectrum and DSA, is to gain access to the spectrum whitespace where the opportunistic transmissions should occur. From the technical and regulatory point of view, there are two main ways of determining the spectrum whitespace.

- Spectrum sensing. Spectrum sensing refers to the proactive activity in the DSA system to capture its ambient spectrum activities. It requires the sensor to have the receiver equipment to continuously measure ambient signal levels across frequency and time. For instance, an SU may be allowed to transmit in a certain channel where the spectrum sensing result reports no incumbent activity in that channel, and the SU may be required to terminate transmission where the spectrum sensing reveals PU activities in that channel. Multiple sensors located at different sites may collaborate to form more reliable determinations of spectrum availability for a given time and location [41–43].

- Geo-location database. This method refers to a centralized solution that requires the user to report its operating location and query the spectrum availability from the database. The geolocation database solution has been practically adopted for

the reasons that the SUs are not required to deploy spectrum sensing hardware and that a centralized solution is beneficial for regulatory purposes [26, 44]. The database should be maintained by the database operators certified by the regulatory body and can be established by various models. The OfCom TV whitespace model requires the PUs to register their activity status, frequency usage, location, and transmission power to the database, and the database employs a predefined propagation model based on geographical information to determine whether the requested frequency at a certain location is immediately accessible or not, as well as the maximum allowed equivalent isometry radiated power (EIRP) of the intended SU's transmission [5, 6]. It is noted that the accuracy of the propagation model largely depends on the geographical environment which can be highly dynamic and complex especially in densely built areas, hence a simplistic and static geographical model can not be relied on for accurate propagation modeling. Given that, to reliably protect public TV services of PUs, the geolocation database has to give very conservative spectrum availability to SUs, and the total throughput of the entire wideband spectrum is far from optimized [45]. Moreover, this model does not require SUs to register their activities to the database, and consequently, interference among multiple SUs accessing the same spectrum hole is likely when they operate in close proximity. The FCC's CBRS model consults spectrum sensing facilities to establish the geolocation database [29]. Specifically, the database, termed as spectrum access system (SAS), may incorporate information from a sensor network that detects transmissions from Department of Defense radar systems and transmits that information to the SAS [9]. Coexisting fixed satellite services are specially protected by registering the ground station locations with the centralized database.

2.2 State-of-the-Art of Wideband Spectrum Sensing

Accurate information of spectrum availability relies on proactive spectrum sensing as the core procedure of DSA. Whether or not the SU performs the spectrum sensing itself or relies on the database that fuses results from spectrum sensing infrastructures, spectrum sensing allows the SU to reliably detect the spectrum holes and opportunistically utilize these frequency resources [23]. Consider the spectrum where multiple but sporadic frequency slices are occupied, and the spectrum of interest should be modeled as a multiband signal spanning wideband frequency ranges. In terms of radio propagation, 'wideband' means the total bandwidth to sense exceeds the coherent bandwidth of each transmission, as opposed to narrowband' which implies the coherence bandwidth of transmission is wider than the whole bandwidth of interest [46].

2.2.1 Conventional Techniques

Wideband spectrum sensing systems can be achieved by conventional radio receiver architectures which can be generally classified into non-real-time and real-time techniques.

A widely adopted approach for non-real-time wideband spectrum sensing is sequential scanning, which incorporates a tunable narrow bandpass filter (BPF) at its radio-frequency (RF) front-end to cover all frequency bands sequentially to detect the existence of incumbent transmissions [47–49]. The tuning range and bandwidth of each BPF selection should be predefined. In Fig. 2.1, the architecture of the sequentially scanning approach is presented, and the corresponding sensing pattern in frequency and time is given in the lower part in Fig. 2.2. It can be seen that this architecture only requires

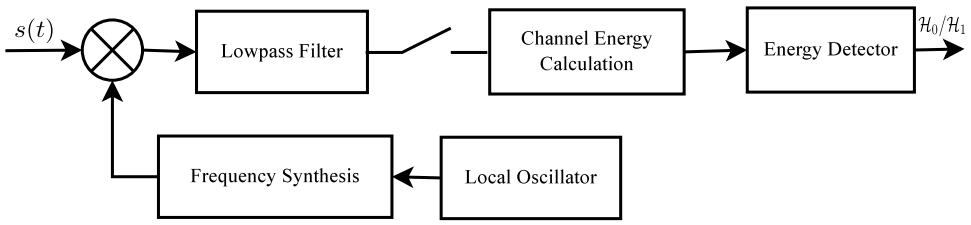


Figure 2.1: Architecture of the sequential scanning scheme of wideband spectrum sensing.

one analog-to-digital converter (ADC) to sample at a rate over the Nyquist rate of each narrow band. However, due to the small real-time bandwidth and the sequential fashion of spectrum scan, the sequential scanning architecture can potentially miss highly dynamic and short lived signals, for example, fast frequency-hopping signals. This can be referred to as the 'blind spots' of sensing in the time-frequency pattern in Fig. 2.2. Consequently the spectrum sensing accuracy mainly depends on the real-time bandwidth of the receiver chain and how fast the real-time band switches. For the application of wideband spectrum sensing and DSA, shortcomings of long scanning time and the resultant missed detection opportunities are fundamental limitations. The scanning speed for such sequential scanning scheme is a critical parameter - the higher of the speed, the shorter duration of the spectrum blind spot. It is reported that the leading spectrum

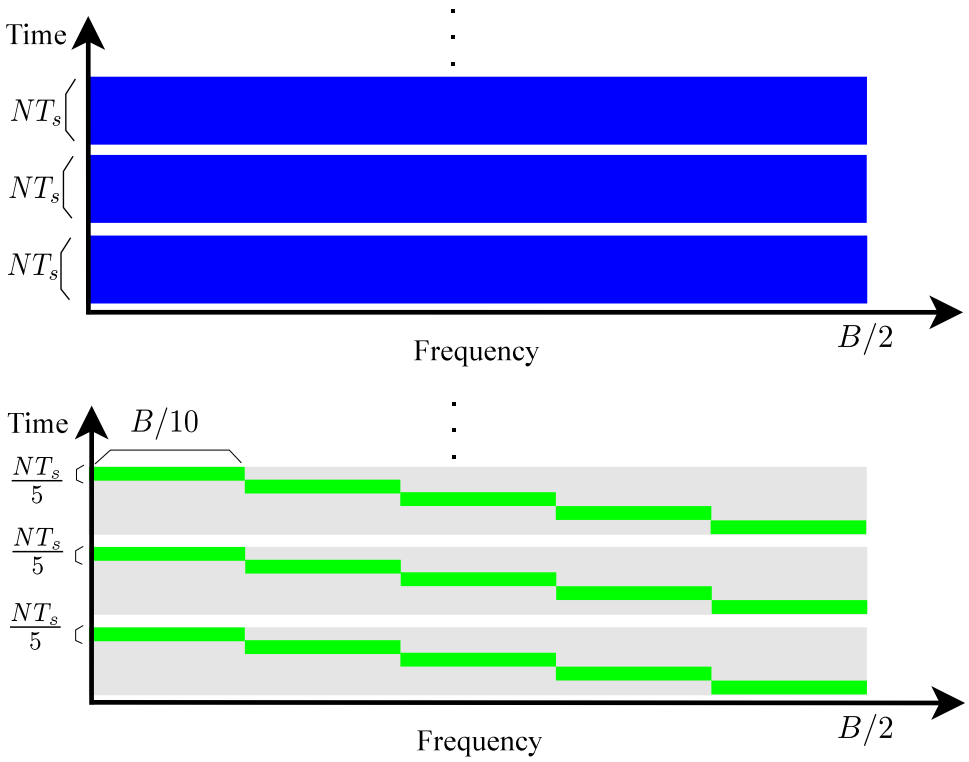


Figure 2.2: Pictorial scanning pattern illustration of wideband real-time spectrum sensing (upper) and sequential scanning (lower). Both schemes provide sensing bandwidth of $\left[-\frac{B}{2}\text{ (Hz)}, \frac{B}{2}\text{ (Hz)}\right]$ and spectral resolution B/N (Hz). The sequential scanning has the instantaneous bandwidth of 1/5 of the whole sensing bandwidth, and the sampling rate of 1/5 of the real-time sampling rate T_s. Those grey portions are the 'blind spots' of the radio activities.

Table 2-A: Summary of RFeye hardware and IEEE 802.22 (FCC TV Whitespace) key specification employing sequential scanning schemes.

Hardware/ Scheme	**Instantaneous Bandwidth**	**Full Bandwidth**	**Scanning Speed**
RFeye 20-6 [51]	20 MHz	10 MHz to 6 GHz	55.6 ms/GHz
RFeye 100-18 [50]	100 MHz	9 kHz to 18 GHz	2.6 ms/GHz
802.22 fast sensing [52]	8 MHz typical	54 MHz to 862 MHz	16.7 ms/GHz
802.22 fine sensing [52]	8 MHz typical	54 MHz to 862 MHz	416.7 ms/GHz

sensor series from RFeye has the scanning rate ranging from 18 GHz per second [50] to 390 GHz per second [51], equivalent to about 55.6 ms per GHz and 2.6 ms per GHz. IEEE 802.22 specification for TV whitespace dictates the sensing time should be less than 1ms per channel in the fast sensing mode and 25ms in the fine sensing mode to ensure quality of service specifications [52], which translate to 16.7 ms per GHz and 416.7 ms per GHz respectively considering a typical channel bandwidth of 6 MHz.

Real-time wideband spectrum sensing removes the inherent problem of missed sensing opportunities but requires more complex hardware architecture or critical hardware performance [53, 54]. One conventional real-time wideband spectrum sensing technique is referred to as the multiband joint detection [43], where a single high-speed ADC is required to sample over the Nyquist rate. The architecture of multiband joint detection is illustrated in Fig. 2.3 and its sensing pattern can correspond to the upper subfigure in Fig. 2.2. After the serial-to-parallel conversion, the time-domain samples are transformed to spectrum domain by a fast Fourier transform block, and the channel

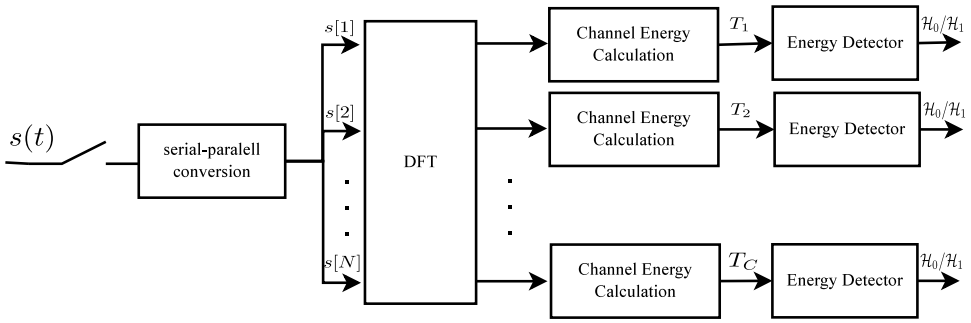


Figure 2.3: Architecture of the multiband joint detection scheme of wideband spectrum sensing.

occupancy can be simultaneously determined for each narrowband by grouping the frequency bins corresponding to each channel and performing energy detection through binary hypothesis test. The high-speed sampling rate of the ADC is the critical limiting factor, as the total bandwidth of the spectrum in applications may span multiple hundreds of megahertz or even multiple gigahertz. Another conventional solution is the filterbank algorithm, as shown in the schematic Fig. 2.4, where it uses multiple ADCs each sampling at a low rate as the sequentially scanning solution but needs multiple dedicated receiver chains, each of which tuned at the central frequency corresponding to each narrow band. Compared with the sequential scanning solution, this filterbank architecture achieves full real-time bandwidth and hence mitigates the problem of long scanning time and blind spots of sensing. However, the expense is the requirement for a great number of RF front-end components and ADCs, significantly increasing the implementation costs and reduce the practicality of realization.

2.2.2 Sub-Nyquist Techniques

In many sensing applications nowadays, including digital imaging, radar, channel estimation, and spectrum sensing, the ever-increasing demand for the scale and accuracy of the sensed data requires the Nyquist rate so high that the excessive number of samples makes compression a necessity prior to storage or transmission. On the other hand, increasing the sampling rate is rather costly, either due to the sensor itself being expen-

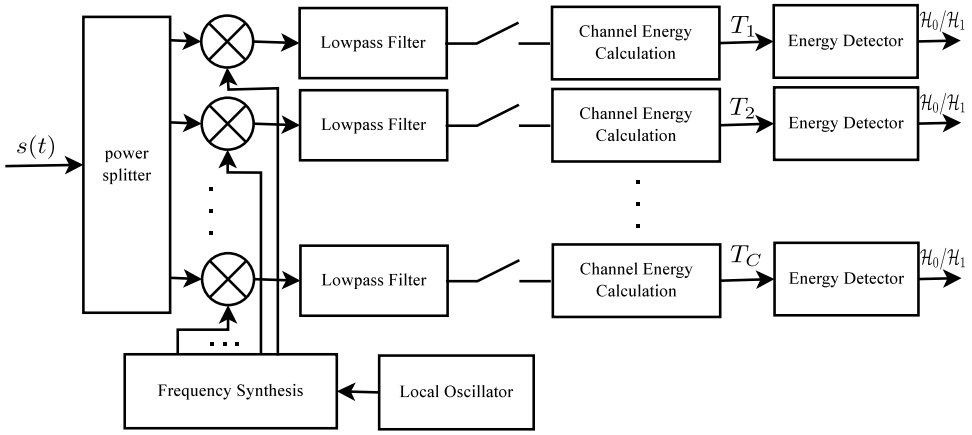


Figure 2.4: Architecture of the filterbank scheme of wideband spectrum sensing.

sive or the measurement process being costly, causing significant degradation in other aspects, for example, energy efficiency [55].

There have long been quests for sub-Nyquist signal acquisition and reconstruction to alleviate the ever-increasing bandwidth of the signals that the communication and electronic systems deal with. As early proposals focus on non-uniform sampling patterns and interpolation reconstruction with prior knowledge of the signals spectrum support, the possibilities of spectrum-blind sub-Nyquist sampling and signal reconstruction were not brought to extensive attention of the researchers until the birth of the CS theory, pioneered by Candes, Tao [56], and Donoho [57] from the 1990s. There are high possibilities that the signals of interest are not arbitrary [55]. Instead, in many cases signals dealt with are 'compressible' in the context of CS - in other words, they can be presented by a set of sparse or nearly sparse non-zero elements, either by itself or under some linear transformation. CS exploits this sparse feature of signals as the precondition, which can achieve a sampling rate significantly lower than the Nyquist rate. The details of compressive sensing models are introduced in Section 2.4.1. Extensive work on compressive sensing theories has revealed a series of performance guarantees of exact or robust recovery from compressed (sub-Nyquist-rate) signals, which depends on collectively the sparsity level of the unknown signal, the sampling structure, and the dimension of the unknown sparse signal and the compressed measurements. In spectrum sensing applications employing compressive sensing, attempts have been seen to adaptively adjust the number of collected compressed measurements to approach the maximum possible compression to ensure reliable recovery of sensing results [58–60]. More detailed review of such performance guarantees are present in Section 2.4.2 and 2.4.3.

It is noted that for conventional real-time wideband spectrum sensing techniques, the Nyquist rate dictates the total sampling rate necessary to represent the wideband signal, whether a single high-rate ADC or parallel low-rate ADCs are used. To alleviate the critical requirement for the sampling rate and hardware complexity of real-time wideband spectrum sensing, the recent theory of compressive sensing has drawn great attention to

achieve real-time signal reconstruction using sub-Nyquist sampling [13, 61].

Employing compressive sensing techniques in wideband spectrum sensing applications to alleviate high sampling rate has been seen in literature in recent years [15, 42, 61–65]. Compressive sensing is enabled in spectrum sensing based on the assumption that the signal to be sensed has sparse representations related to its frequency domain, most commonly in the original spectrum [13, 15, 16, 53, 61, 63, 65], or alternatively the average power spectrum [62, 66, 67] or the cyclic power spectrum [68, 69]. To enhance detection performance of the spectrum sensing by exploiting the spatial diversity of radio propagation, compressive spectrum sensing (CSS) in the cooperative scenario has also been proposed and studied in both centralized [13–16] and decentralized [64, 70] sensor networks. Cooperative spectrum sensing strategies may congregate measurements in their compressive sensing models from individual sensing nodes, forming a low-rank matrix completion [13, 15] or joint 1-norm sparse recovery problem [16] to obtain consensus sensing results from multiple sensing measurements.

The difference between architectures of real-time conventional non-compressive spectrum sensing and mainly lies in the introduction of the sub-Nyquist sampler hardware and the CS recovery algorithm, in substitution of the combination of a wideband Nyquist-rate ADC and a Fourier transform block. Specifically, the wideband CSS system is implemented by the following steps,

- sub-Nyquist-rate measurement acquisition;
- reliable signal reconstruction via sub-Nyquist-rate measurements;
- detection of incumbent radios from the recovered frequency-domain signal.

Sub-Nyquist Sampling Architecture The sub-Nyquist sampling hardware in a CSS system converts the analog signal of interest into discrete-time sub-Nyquist-rate compressed samples, and the sampling pattern of it determines the sampling matrix of the CS model and hence the performance of the sparse recovery. There is a considerable

amount of work from the design, modeling to the practical sub-Nyquist samplers for wideband sparse signals. However, all of the practical sub-Nyquist samplers for wideband signals have inherent shortcomings in terms of realization complexities, constraints on the CS mathematical model, and hardware constraints. An analog-to-information converter, also referred to as a random demodulator, has been proposed in [71, 72] as the earliest practical solution of data acquisition in CSS. It employs a front-end mixer to make the input signal scrambled by a pseudorandom sequence, followed by an integrator and a sampling device to dump the discrete-time measurements, which can be shown to form a single-measurement-vector (SMV) CS model with a block-diagonal sensing matrix [71, 73]. The main drawback of the random demodulator is that it often needs a large signal length in the SMV model to provide sufficient spectral resolution, and as a result, the random demodulator is commonly proposed for sensing multitone signals [65, 73]. Real-world implementations of the random demodulator are present in the literature, and most wideband realizations are low-level designs of integrated circuits on semiconductors [74–76], instead of using off-the-shelf components due to the novel structure of the random demodulator, especially the requirements for the mixer-and-integrator (filter) device, referred to as random demodulator pre-integrator (RDPI) [74, 76, 77]. Practical topics of the design and the performance evaluation of the random demodulator have been extensively studied by Becker and Baraniuk in [77, 78]. The multicoset sampler, however, employs the conventional architecture of non-uniform sampling [79] and uses parallel low-rate ADCs with distinctive signal delays to construct a multiple-measurement-vector (MMV) model of CS with a structured sampling matrix dependent on these time delays. Such an architecture has been brought to attention to sub-Nyquist sampling and drawn particular interests due to its simple hardware structure and suitability for low-complexity joint sparse recovery algorithms for its MMV model [16, 61, 66]. However, the major disadvantage of the multicoset sampler is that the low-rate ADCs need to have a large front-end bandwidth no less than the signal's Nyquist frequency [65]. More recently, Mishali et. al. [65] proposed the architecture of the modulated wideband converter (MWC), which somewhat combines the parallel

low-rate ADCs from the multicoset sampler and the front-end mixer configuration from the random demodulator. It effectively removes the problems with both abovementioned architectures - it has an MMV model of CS and only requires a low front-end bandwidth equal to each ADC's sub-Nyquist sampling rate [65, 80]. Moreover, the sampling matrix of the MWC model of CS can potentially be fully random, determined by the scrambling sequence of all branches, implying superior sparse recovery performance. The shortcoming of MWC lies in technical aspects such as the relatively high system complexity and synchronization of Nyquist-rate scrambling sequences. A 6 GHz real-world demo of the MWC system has been established, and the particular technical challenges, such as the design of synchronized multi-gigahertz scrambling sequences and high-throughput data acquisition solutions, are illustrated in which shed light in future implementation efforts of practical CSS architectures [81, 82]. More detailed introductions of the architecture and modeling of the three sub-Nyquist sampling strategies are presented in Section 2.5. Moreover, it is of particular interest to understand the implications of the sub-Nyquist sampling on the radio receiver's performance, i.e. the distortions of the received (or recovered) signals due to imperfections in the sampling process. Recent studies [83, 84] has shown that with the presence of additive noise, one can still largely expect the '3 dB per octave sampling frequency' gain in the signal's signal-to-noise ratio (SNR), which implies a negative SNR gain in the context of sub-Nyquist sampling. Also, it reveals that the '6 dB per bit' quantization SNR improvement for conventional samplers still largely applies to a sub-Nyquist sampler, but potentially further improvements with signals of high crest factor due to the nature of compressive sampling patterns [83]. A review of noise performance on the sub-Nyquist samplers is detailed in Section 2.5.4.

Recovery Algorithms The recovery of the unknown sparse signal relating to the second step above is a major problem in CSS systems. Tractable solutions to such a sparse recovery problem can be found by the relaxation of the original 0-norm to a v-norm objective ($0 < v \leq 1$), where non-convex ($0 < v < 1$) and convex optimization ($v = 1$) algorithms can be applied [55]. Especially, the 1-norm relaxation has drawn more atten-

tion in terms of both the research of its performance and application due to the convexity of the objective. The algorithms for it are available in various convex optimization toolboxes, such as SPGL1 [85] and FPC [86]. Non-convex optimization ($0 < v < 1$) has also been studied [87], with efforts seen in CSS applications to reduce the complexity of the optimization routines [53, 88, 89]. Detailed introduction to v-norm ($0 < v \leq 1$) optimization algorithms is given in Section 2.4.4.1 and 2.4.4.2. Greedy algorithms, as effective strategies to pursue a sparse solution, have inferior performance in terms of stable and exact recovery of general sparse signals compared with optimization methods. However, their main advantage is the simplicity of formulation and computational complexity, which is particularly crucial in spectrum sensing applications where fast sensing speed is highly desired. Technically, the greedy algorithms for sparse recovery are not optimization algorithms since they are not driven by an overall objective. Instead, they build a solution – or the support of the solution – part by part, yet in many cases, the solution can be guaranteed the sparsest solution [55]. In contrast to early work in CSS which mainly proposed the use of optimization techniques [13–15, 63], recent developments in CSS tend to pay more attention to greedy algorithms for its lower complexity to tackle the challenge of the real-time implementation of CSS [16, 65, 80]. A detailed review of greedy algorithms is present in Section 2.4.4.3. However, an inherent problem of greedy algorithms is that the sparsity level of the unknown signal should be a priori known, and hence a sparsity estimation scheme is necessary [58, 59, 90]. A more recent sparse recovery algorithm called sparse Bayesian learning (SBL) [91] has also been seen in the CSS literature [92, 93], where it reportedly has less recovery errors than v-norm optimization counterparts ($1 < v \leq 1$) but still forms a non-convex problem and has relatively high computational complexity [94]. Introduction to the algorithm routines for SBL is presented in Section 2.4.4.4. In CSS applications, the use of readily known data, such as the information from geo-location database infrastructures, to assist the sparse recovery algorithm have been proposed [14, 88, 95]. Specifically, such assisting data can provide initial points for some iterative sparse recovery routines, in order to reduce the computational complexity, as well as to effectively reduce the convergence

error in the recovered signal for non-convex optimization algorithms [88, 95].

Incumbent Radio Detection The classic and most commonly adopted approach in the literature of determining the spectrum availability is Neyman-Pearson energy detection [96–98]. Specific models can be derived for different channel models, such as line-of-sight, Rayleigh or Rician channel [99, 100]. Due to the difficulty of obtaining prior information of the incumbent signal, a common practical approach to determine the threshold is by the constant-false-alarm-rate (CFAR) criterion, where the threshold only corresponds to the null hypothesis relating to the noise. The energy threshold can be determined by inversely finding the integral bound once the target false alarm rate is given. Some proposals have been noted to employ a special CS model of the cross-correlations of the stationary time-domain signal and recover the sparse power spectrum directly [62, 66, 67]. However, the stationarity assumption of the incumbent signal may not be always satisfied, especially when the active signal tends to be short-lived. Another classic detection technique of wideband spectrum sensing is cyclic feature detection [101–104]. It employs the cyclostationarity of the incumbent signal transmission and reveals the two-dimensional cyclic power spectrum. Based on the different cyclostationary features in the cyclic spectrum, signal activities and their modulation types can also be recognized. Another advantage over energy detection is that the cyclic feature detection has more robustness against noise as the cyclic features of incumbent signals are generally cyclostationary but the noise (generally assumed stationary) has no cyclostationarity [68, 101]. To realize direct cyclic power spectrum reconstruction from compressed samples, it is proposed to design some specialized CS models on the cross-correlations of the time-domain signal, and the coefficients of the cyclic spectrum can be recovered as the unknown sparse signal in the CS model [68, 69]. Although such cyclic power spectrum recovery strategies may additionally employ the sparsity of cyclic frequency domain to further reduce the sampling rate [69], one major concern of the CS-based cyclic feature detection is the high computational complexity of recovering the two-dimensional spectrum. Another inherent shortcoming is that it generally requires

a long observation window in the signal acquisition to produce good estimations of the cyclic power spectrum, which is in contrast to the fast sensing requirement of spectrum sensing applications.

2.3 Sampling Theories

Here, the classic lowpass and bandpass sampling theories are presented. Moreover, a recently derived lower-bound of sampling rate for spectrum-blind bandpass signals has been noted, which can serve as a sufficient condition on the sampling rate for sub-Nyquist wideband spectrum sensing.

2.3.1 Lowpass Sampling

The lowpass sampling theorem, also known as the Nyquist-Shannon sampling theorem, has been long perceived as the fundamental theory in signal discretization and discrete-time signal processing [105]. Consider a complex-valued continuous-time signal, which a continuous function of time t, $s(t) \in \mathbb{C}$, $\forall t \in \mathbb{R}$, whose frequency presentation is given by Fourier transform and is assumed to exist as such

$$sX(f) := \int_{-\infty}^{+\infty} s(t) \exp(-j2\pi ft)\mathrm{d}t, \tag{2.1}$$

where the function $X(f)$ of $f \in \mathbb{R}$ is piecewise continuous. For bandlimited signals $s(t) \in \mathcal{M}_B$ where the set is formally expressed by $\mathcal{M}_B := \{s(t)|X(f) = 0, f \notin [-B/2 \text{ (Hz)}, B/2 \text{ (Hz)}]\}$, a uniform sampling sequence at the rate of f_s generating the discrete series $\{s[n] = s(n/f_s), n \in R\}$ can perfectly recover $s(t)$ if

$$f_s \geq B := f_{\text{NYQ}} \tag{2.2}$$

where f_{NYQ} is namely the Nyquist rate. Specifically, the continuous-time signal $s(t)$ can be reconstructed by the following interpolation expression,

$$s(t) = \sum_{n=-\infty}^{\infty} s[n] \operatorname{sinc}\left(f_s t - n/f_s\right), \tag{2.3}$$

known as the Whittaker-Shannon formula [105]. It is noted that the interpolation function is the impulse response of an ideal brick-wall lowpass filter for the passband of $f \in [-B/2\ (\mathrm{Hz}), B/2\ (\mathrm{Hz})]$.

2.3.2 Bandpass Sampling

Moreover, the lowpass sampling theorem has been generalized to the bandpass form by Landau in 1960s, for any wideband signal bandlimited within $f \in [-B/2\ (\mathrm{Hz}), B/2\ (\mathrm{Hz})]$ with known occupied passbands forming a set $\mathcal{T} \subseteq [-B/2\ (\mathrm{Hz}), B/2\ (\mathrm{Hz})]$, formally expressed by the set $M_{\mathcal{T}} = \{s(t) | X(f) = 0, f \notin \mathcal{T}\}$. Landau [106] has proved that, for an average sampling rate f_s, there exists a sampling pattern which can form a stable sequence of discrete signals which leads to perfect reconstruction of $s(t)$, if

$$f_s \geq \lambda(\mathcal{T}) := f_{\mathrm{LAN}}, \tag{2.4}$$

where $\lambda(\mathcal{T})$ denotes the Lebesgue measure of the set $\mathcal{T}$, in other words, the sum of actual occupied bandwidth of each passband of the signal $s(t)$; and f_{LAN} denotes the Landau's rate [61].

Compared with the lowpass sampling theorem, Landau's sampling theorem implies the minimum average sampling rate can be largely saved if the actual spectrum occupancy is sparse. It should be noted that Landau's sampling theorem requires the actual spectrum support T to be a priori known so that it is possible to have a proper sampling pattern with a rate above Landau's rate to be designed. The Landau's rate, however, does not reveal how to construct the sampling pattern or dictate the uniqueness of the sampling pattern. Research efforts of practical solutions towards Landau's rate have

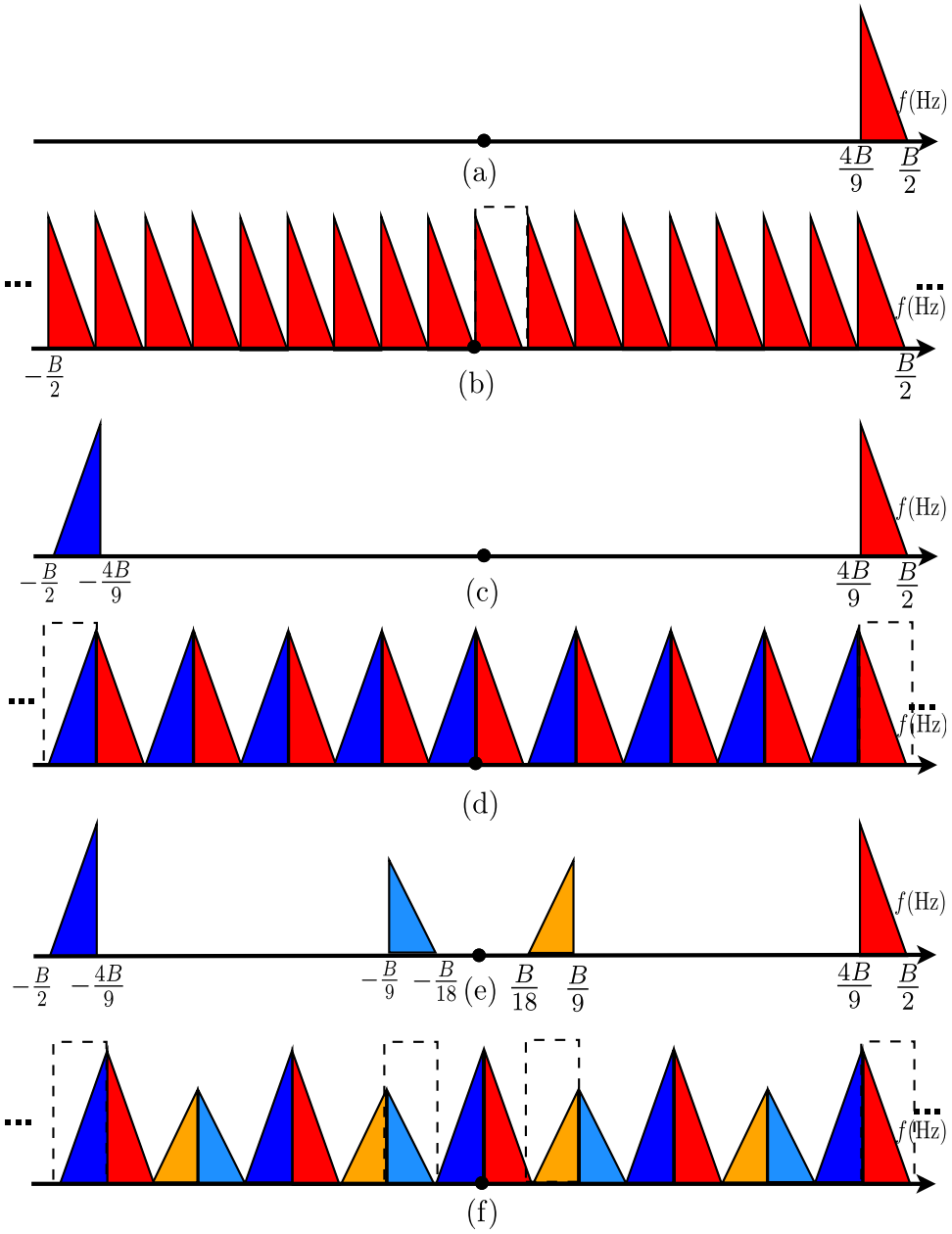


Figure 2.5: Illustrations of the 'spectrum-folding' strategy of achieving Landau's rate. **(a)** the spectrum of a single pass-band (complex) signal where the spectral support is $\left[\frac{4B}{9}\text{ (Hz)}, \frac{B}{2}\text{ (Hz)}\right]$; **(b)** the resulting spectrum of the discrete-time version of the signal in (a) using $f_s = \frac{B}{18}$ (Hz) with the anti-aliasing band pass filter for reconstruction; **(c)** the spectrum of a (real) signal where the pass band is $f \in \left[\frac{4B}{9}\text{ (Hz)}, \frac{B}{2}\text{ (Hz)}\right]$ with its conjugate symmetric band in the negative frequencies; **(d)** the resulting spectrum of the discrete version of the signal in (c) using $f_s = \frac{B}{9}$ (Hz) with the anti-aliasing band pass filters for reconstruction;**(e)** the spectrum of (real) signal where the pass bands are $f \in \left[\frac{4B}{9}\text{ (Hz)}, \frac{B}{2}\text{ (Hz)}\right] \cup \left[\frac{B}{18}\text{ (Hz)}, \frac{B}{9}\text{ (Hz)}\right]$ with its conjugate symmetric bands in the negative frequencies; **(f)** the resulting spectrum of sampling the signal in (e) using $f_s = \frac{2B}{9}$ (Hz) with the anti-aliasing band pass filters for reconstruction.

often been dedicated to a specific class of multiband signals and certain types of practical uniform or non-uniform sampling schemes [79, 107–109].

One commonly used strategy of achieving Laudau's rate for bandpass signals can be summarized by designing an sampling rate $f_s = B/N_0$ (Hz) ($N_0 > 1, N_0 \in \mathbb{Z}$) to fold the pass band components in the frequency domain to other frequencies without causing the aliasing of the pass band components themselves [110]. For instance one can consider a very simple case where the signal occupies a single band, $f \in \left[-\frac{B}{2} + \frac{N_2}{N_1}\frac{B}{2}\ (\text{Hz}), -\frac{B}{2} + \frac{N_2+1}{N_1}\frac{B}{2}\ (\text{Hz})\right]$, for some $N_1, N_2 \in \mathbb{N}$ satisfying $0 < N_2 \leq 2N_1 - 1$. An example for this setting is shown in Fig. 2.5 (a), where $N_1 = 9$ and $N_2 = 17$. One can choose a sampling rate of $f_s = f_{\text{LAN}} = \frac{B}{2N_1} = \frac{B}{18}$ (Hz) to under-sample the continuous-time signal without aliasing the pass band components. The under-sampled signal is shown in the frequency domain in Fig. 2.5 (b). Next, it is straightforward from the Whittaker-Shannon theorem to prove that continuous-time pass band signal $s(t)$ can be interpolated by $\{s[n] = s(n/f_s)\}$ using the corresponding impulse response of an ideal brick-wall bandpass filter, as outlined in Fig. 2.5 (b), for the pass band of $f \in \left[-\frac{B}{2} + \frac{N_2}{N_1}\frac{B}{2}\ (\text{Hz}), -\frac{B}{2} + \frac{N_2+1}{N_1}\frac{B}{2}\ (\text{Hz})\right]$, specifically

$$s(t) = \sum_{n=-\infty}^{\infty} s[n] \cdot \frac{\exp(j2\pi(-B/2 + \frac{N_2 B}{2N_1})t) - \exp(j2\pi(-B/2 + \frac{(N_2+1)B}{2N_1})t)}{j2\pi t}, \qquad (2.5)$$

where $N_1 = 9$ and $N_2 = 17$. It should be noted that the example in Fig. 2.5 (a) and (b) are based on complex time-domain signal $s(t)$, inline with previous discussions of low-pass sampling theorem. In practice, the same strategy can be used for real-valued signals, where the spectrum is conjugate symmetric. Two more examples are illustrated for real-valued signals, in Fig. 2.5 (c) and (e), for one and two pairs of conjugate symmetric pass bands respectively. One can see the Landau's rate for both cases as $f_{\text{LAN}} = \frac{B}{N_1} = \frac{B}{9}$ (Hz) and $f_{\text{LAN}} = \frac{2B}{N_1} = \frac{2B}{9}$ (Hz) respectively, and using the sampling rate $f_s = f_{\text{LAN}}$ can make the under-sampled signals fold in frequency domain as shown in Fig. 2.5 (d) and (f) respectively. Finally, the original signals can be reconstructed by

a similar process as (2.5), but using two and four ideal bandpass filters, as outlined in Fig. 2.5 (d) and (f) respectively.

A more general signal reconstruction system achieving Landau's rate can be achieved by a type of practical non-uniform sampling pattern referred to as multicoset sampling [79]. In [79], it is proven that for a signal with a known support of multiple equally spaced bands, that is, for a given number of division $2N_1$ of $[-B/2 \text{ (Hz)}, B/2 \text{ (Hz)}]$, any signal belonging to the set $\mathcal{T} = \cup_{n_2 \in \mathcal{N}_2} \left[-\frac{B}{2} + \frac{n_2}{N_1}\frac{B}{2} \text{ (Hz)}, -\frac{B}{2} + \frac{n_2+1}{N_1}\frac{B}{2} \text{ (Hz)}\right)$, where $\mathcal{N}_2$ is a subset of $\{n_2 | n_2 \in \mathbb{N} \text{ s.t. } 0 < n_2 \leq 2N_1 - 1\}$, can be perfectly reconstructed using samples from a multicoset sampler with an average sampling rate no less than the Landau's rate of $f_{\text{LAN}} = \frac{B|\mathcal{N}_2|}{2N_1}$ (Hz). With the knowledge of set $\mathcal{N}_2$, i.e. the actual spectrum occupancy, the requirements for the specific non-uniform sampling pattern in the context of multicoset sampling and the corresponding interpolation formula can be determined in order to exactly reconstruct the continuous-time signal $s(t)$.

2.3.3 Spectrum-Blind Sampling

In the background of wideband spectrum sensing, one would wish to reconstruct the original multiband signal $s(t)$ without the knowledge of the spectrum occupancy, termed as spectrum-blind constraints, which is more stringent than the known-spectrum condition in the Landau's sampling theorem. That said, one would wonder if a similar lower-bound of sampling rate for exact reconstruction exists in this general case of interest. The result has been recently formalized in [61] as a generalization of Landau's sampling theorem. The specific term of spectrum-blind is defined as such that the information about the band locations is not used while acquiring the samples and neither can it be used in the reconstruction process. It is stated that for any signal in the set $\mathcal{M}_K = \{s(t) | \lambda(\mathcal{T}) \leq \frac{K}{B}, 0 < K < 1 \text{ and } X(f) = 0, \forall f \notin [-B/2 \text{ (Hz)}, B/2 \text{ (Hz)}]\}$ with $\mathcal{T}$ being set of the occupied frequencies, there exists a sampling pattern with an average sampling rate of f_s (Hz) that the sampling pattern can form a stable sequence of discrete

signal which leads to perfect reconstruction of $s(t)$, if

$$f_s \geq \max\left(2K/B \text{ (Hz)}, B \text{ (Hz)}\right) = \max\left(2f_{\text{LAN}}, f_{\text{NYQ}}\right). \tag{2.6}$$

This theorem implies that in order to perfectly reconstruct any bandlimited multiband signals, theoretically, one has to pay the expense of sampling at least at twice the Landau's rate or the Nyquist rate, due to the lack of knowledge of the locations of the signal's spectrum occupancy.

2.3.4 Remark

It is worth remarking that all three sampling theorems mentioned in this section are sufficient conditions. In other words, their converse proposition is not true - it is not valid to say, taking the inverse of the lowpass sampling theorem for example, that a signal $s(t) \in \mathcal{M}_B$ cannot be perfectly reconstructed using a uniform sampling rate less than $f_{\text{NYQ}} = B$ (Hz). It is easy to find a variety of signals, for example $s(t) \in \mathcal{M}_{\mathcal{T}} \subset \mathcal{M}_B$, that can be perfectly reconstructed using a uniform rate of $f_{\text{LAN}} = B/2N_1$ (Hz) $< f_{\text{NYQ}}$, as detailed in Section 2.3.2. The same sense applies to minimum rates in the bandpass and spectrum-blind sampling theories.

Moreover, these theorems are all established using the conventional paradigm of signal reconstruction, that is by the linear combinations of basis functions (i.e. interpolation functions) in certain functional spaces. The recent research topic of compressive sensing (CS) turns to another paradigm of signal reconstruction and instead proposes to apply optimization techniques on the sub-Nyquist-rate samples. Comments on the sufficient sub-Nyquist sampling rate leading to exact or stable reconstruction are given in a completely different context, seen in Section 2.4.2. It is particularly interesting to see the coincidence of the minimum rate of the spectrum-blind sampling in the setting of CS with Theorem 2.4.5 and 2.4.6 in Section 2.3.3.

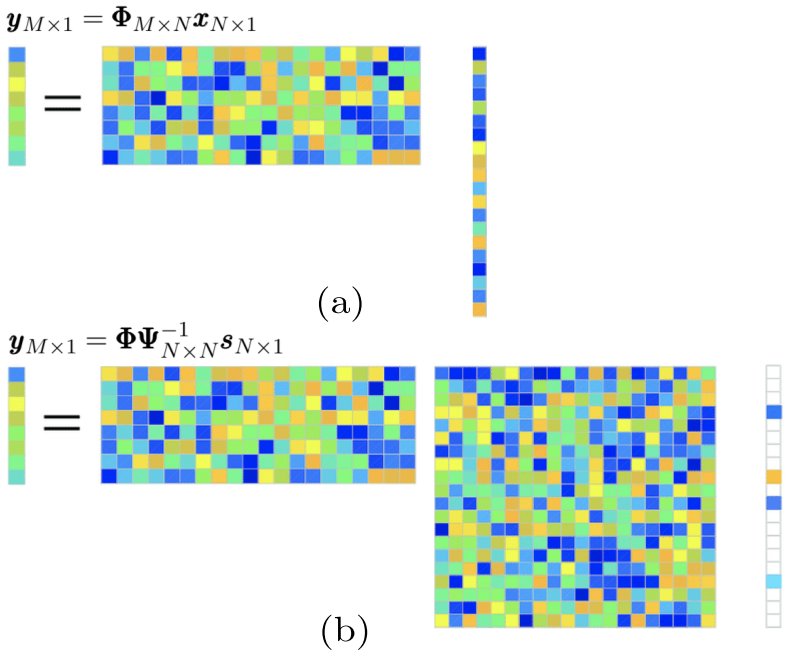


Figure 2.6: CS mathematical model. **(a)** Compressive sampling of a compressible signal; **(b)** Compressive sampling with the sparsifying matrix and the sparse signal.

2.4 Compressive Sensing

2.4.1 Mathematical Model

The sub-Nyquist sampling model in the context of CS is presented by a linear system, writing

$$\boldsymbol{y}_{M\times 1} = \boldsymbol{\Phi}_{M\times N}\boldsymbol{x}_{N\times 1} + \boldsymbol{b}_{M\times 1} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}_{N\times N}\boldsymbol{s}_{N\times 1} + \boldsymbol{b}, \tag{2.7}$$

where $\boldsymbol{y}$ is a vector with sub-Nyquist-rate measurements; $\boldsymbol{\Phi}$ is the sampling matrix modeling the sampling behaviors of the practical sampling hardware; $\boldsymbol{x}$ is the original Nyquist-rate presentation of the signal to be sensed; $\boldsymbol{\Psi}$ is the sparsifying matrix which represents the linear transformation towards the sparse presentation $\boldsymbol{s}$, and $\boldsymbol{b}$ is additive noise. The product $\boldsymbol{A} := \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$ is referred to as the sensing matrix. Note it is required that $M < N$ to ensure the measurement vector $\boldsymbol{y}$ is compressed compared to the original signal $\boldsymbol{x}$, thus achieving a sub-Nyquist sampling rate. With the illustration of Fig. 2.6 (a), one can intuitively think each element of the compressed measurement vector being

the inner product between the corresponding row of the sampling matrix and the original signal, where the measurement element may contain a weighed portion of all the elements in the original signal. In Fig. 2.6 (b), the original signal $\boldsymbol{s}$ is transformed into a sparse signal $\boldsymbol{x}$. The sensing matrix, $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$, can be viewed to sample the sparse signal $\boldsymbol{x}$ in the analogous fashion that $\boldsymbol{\Phi}$ samples $\boldsymbol{s}$. Such sampling procedure may be interpreted by a 'democratic' process, where each element of $\boldsymbol{x}$, being zero or not, having its significance to be 'coded' in the elements in the measurement vector $\boldsymbol{y}$ [111].

Consider a noiseless case $\boldsymbol{b} = \boldsymbol{0}$, the linear system (2.7) are under-determined, meaning the arbitrary interested signal $\boldsymbol{x}$ or $\boldsymbol{s}$ can not be uniquely determined by measurements $\boldsymbol{y}$ and sensing matrix $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$. However, with the prior assumption that $\boldsymbol{s} = [s_1, s_2, \cdots, s_N]^T$ is sparse, defined by the support of $\boldsymbol{s}$, i.e. $\text{supp}(\boldsymbol{s}) := \{n | s_n \neq 0, \forall n \in \mathbb{N} \text{ and } 1 \leq n \leq N\} = k << N$, the signal $\boldsymbol{s}$ is proposed to be found by seeking the sparsest solution under this linear system,

$$\boldsymbol{s}^o = \arg\min_{\boldsymbol{s}} ||\boldsymbol{s}||_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{s}, \tag{2.8}$$

where $||\boldsymbol{s}||_0 := |\text{supp}(\boldsymbol{s})|$ denotes the signal's sparsity and it is demonstrated in Section 2.4.2 that the uniqueness of the sparsest solution can be guaranteed under some conditions on the sparse signal $\boldsymbol{s}$ and the sensing matrix $\boldsymbol{A}$.

However, due to the non-continuous zero-norm objective function of (2.8), $\boldsymbol{s}^o$ cannot be found by numerical optimization methods but by combinatorial optimization which is NP-hard [112]. A computationally tractable substitute of problem (2.8) is proposed by using a convex objective $||\boldsymbol{s}||_1$ that mostly approximates $||\boldsymbol{s}||_0$, referred to as convex relaxation, such that

$$\boldsymbol{s}^o = \arg\min_{s} ||\boldsymbol{s}||_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{s}, \tag{2.9}$$

where one-norm is defined by $||\boldsymbol{s}||_1 := |s_1| + |s_1| + \cdots + |s_N|$. It is a fundamental problem in CS that when $\boldsymbol{s}^o = \boldsymbol{s}$ by convex relaxation and optimization, can be uniquely determined by solving the problem (2.9), given various conditions on parameters M, N, k, and the

sensing matrix $\boldsymbol{A}$. Existing answers to this question are referred to as 'exact recovery' and to be reviewed in Section 2.4.2.

For CS to be useful in real-world sensing applications, it must deal with somehow imperfect measurements. A general optimization model considers the presence of additive noise and an inequality constraint, which writes

$$\boldsymbol{s}^o = \arg\min_s ||\boldsymbol{s}||_1 \quad \text{s.t.} \quad ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2 \leq \epsilon, \tag{2.10}$$

where a constant ϵ accounts for the tolerance related to the variance of noise. Another formulation puts the quadratic data fidelity term into the objective and treats the 1-norm as the constraint, forming a constrained least-square problem

$$\boldsymbol{s}^o = \arg\min_s ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2 \quad \text{s.t.} \quad ||\boldsymbol{s}||_1 \leq \xi, \tag{2.11}$$

for which some classic techniques for general quadratic convex optimization could apply, such as least-absolute shrinkage and selection operator (LASSO) algorithms. Here, ξ is the constant controlling the l_1 constraint. Alternatively, instead of a constraint, it is common to replace the constraint by a penalty term in the objective, making the optimization unconstrained and more convenient finding a best solution. The unconstrained version for (2.10) or (2.11) writes

$$\boldsymbol{s}^o = \arg\min_s ||\boldsymbol{s}||_1 + \frac{1}{2\lambda}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2, \tag{2.12}$$

where $\lambda \geq 0$ accounts for the weighing of the noise penalization term. The value of λ should be made large if more noise is allowed in the original signal, and one can consider $\lambda \to \infty$ as the noiseless recovery model approximating (2.9). For some appropriate selections of ϵ, ξ and λ, the solutions of (2.10), (2.11), and (2.12) coincide, and all three forms are often deemed equivalent in this sense. However, the values of the parameter that make the solution coincide for these 'equivalent' problems could not generally be determined a priori.

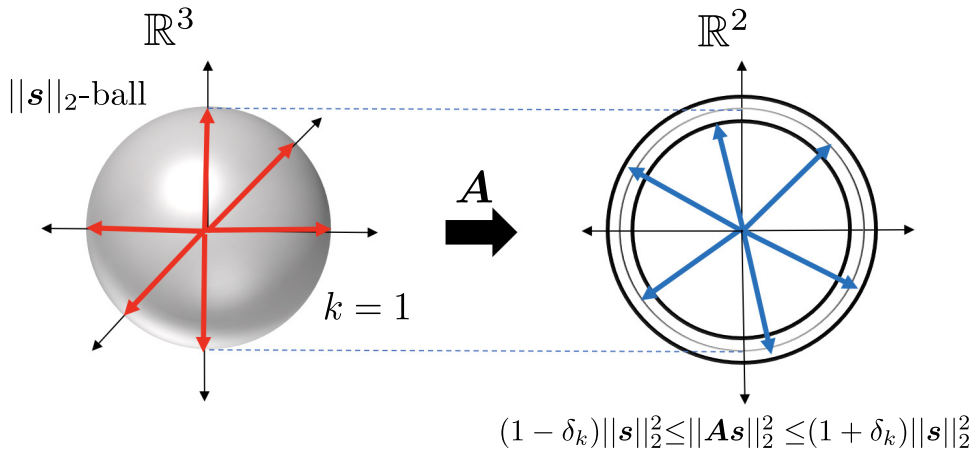


Figure 2.7: Schematic illustration of the RIP with sparsity 1 of $\boldsymbol{A}_{2\times 3}$.

For (2.10), (2.11), and (2.12), the recovery cannot be exact due to the perturbation of the noise, and the conditions of 'robust' recovery satisfying bounded distance between $\boldsymbol{s}^o$ and $\boldsymbol{s}$ is of interest. Further discussions are presented in Section 2.4.2.

2.4.2 Theoretical Guarantees of Performance

There are various theories revealing when the one-norm optimization is capable of finding the most sparse solution among all possible versions of $\boldsymbol{s}$ satisfying (2.7) with or without noise impairments, in other terms, guarantee an exact or stable CS recovery, respectively. These properties are based on various mathematical tools, have different strengths, and they together imply that a large number of sensing matrices, including random matrices and structured random matrices, are applicable for CS and have recovery guarantees. In this section, a few of the most coom theories on the sensing matrices and sparse recovery performance are reviewed.

2.4.2.1 Restricted Isometry Property

Restricted isometry property (RIP) [113] is introduced as the standard tool for evaluating sensing matrices. Specifically, a sensing matrix $\boldsymbol{A}$ is defined to have RIP with isometry constant δ_k, if $0 \leq \delta_k < 1$ is the smallest number to satisfy

$$(1-\delta_k)||\boldsymbol{s}||_2^2 \leq ||\boldsymbol{A}\boldsymbol{s}||_2^2 \leq (1+\delta_k)||\boldsymbol{s}||_2^2 \tag{2.13}$$

with all k-sparse vectors $\boldsymbol{s}$. The expression (2.13) describes the approximate power-preserving (l_2-norm) property of the linear transform $\boldsymbol{A}$ for arbitrary k-sparse vectors. A schematic illustration is shown in Fig. 2.7, where $\boldsymbol{A}_{2\times 3}$ that transforms vectors from $\mathbb{R}^3$ to $\mathbb{R}^2$. The RIP of $\boldsymbol{A}$ with the constant δ_1 in this case can be thought as that all 1-sparse vectors of an arbitrary 2-norm length in $\mathbb{R}^3$ would have their Euclidean length approximately the same in $\mathbb{R}^2$ after the linear transform $\boldsymbol{A}$, where the length scaling is restricted by $\sqrt{1-\delta_1}$ and $\sqrt{1+\delta_1}$.

If $\delta_{2k} < \sqrt{2}-1$ [113], the recovery has been proved to be exact using (2.9), which gives the following necessary condition. Furthermore, such bounds for δ_{2k} also serve as a necessary condition for the stable sparse recovery problem (2.10). Similar conditions have been developed from the original RIP constant $\delta_{2k} < \sqrt{2}-1 \approx 0.4142$ [113] to more recent bounds such as $\delta_{2k} = 0.4652$ [114] and $\delta_{2k} = 0.4931$ [115], and the theorem with the latest bound is presented as follows.

Theorem 2.4.1. *Consider model (2.10). If $\boldsymbol{A}$ has RIP (with $\delta_{2k} < 0.4931$), then the solution $\boldsymbol{s}^o$ satisfies the following with relaxation ϵ set to $||\boldsymbol{b}||_2$*

$$||\boldsymbol{s}-\boldsymbol{s}^o||_1 \leq C_1\sqrt{k}||\boldsymbol{b}||_2 + C_2\sigma_{[k]}(\boldsymbol{s}). \tag{2.14}$$

and

$$||\boldsymbol{s}-\boldsymbol{s}^o||_2 \leq C_3||\boldsymbol{b}||_2 + C_4\sigma_{[k]}(\boldsymbol{s})/\sqrt{k}, \tag{2.15}$$

where C_1, C_2, C_3, C_4 are universal constants only relating to δ_{2k}. The k-sparse approximation error $\sigma_{[k]}(\boldsymbol{s})$ is defined as $||\boldsymbol{s}-\boldsymbol{s}_{[k]}||_2$ where $\boldsymbol{s}_{[k]}$ is the vector obtained by setting all but the k largest (in magnitude) entries of $\boldsymbol{s}$ to zero. [115]

The results claim that as long as the sensing matrix $\boldsymbol{A}$ obeys the RIP, the l_1 recovery converges and is stable against both the measurement noise - $\boldsymbol{b}$ - and the signal noise given in terms of the best k-term approximation error. When no noise appears, the above guarantee simply degenerates to the guarantee of stable recovery, i.e. $\boldsymbol{s}^o = \boldsymbol{s}$. Although

in Theorem 2.4.1 the bounds are given explicitly, the main limitation of RIP guarantees is that the isometry constants cannot be computed in polynomial time for an arbitrary deterministic matrix [107].

2.4.2.2 Mutual Coherence and RIPless

As the isometry constant in the RIP guarantee is not trivial to calculate, to tackle the problem of RIP property and to provide the error bounds for stable recovery and the condition of exact recovery with minimum restrictions on $\boldsymbol{s}$ and $\boldsymbol{A}$, RIPless guarantees based on the definition of mutual coherence have been proposed for any fixed $\boldsymbol{s}$ without the need for evaluating RIP [116]. The coherence of a matrix $\boldsymbol{A}_{M\times N} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_N] \in \mathbb{C}^{M\times N}$ is defined as the maximum coherence between columns [117]:

$$\mu(\boldsymbol{A}) = \max_{i\neq j} \frac{|\boldsymbol{a}_i^H \boldsymbol{a}_j|}{||\boldsymbol{a}_i||_2||\boldsymbol{a}_j||_2}, \qquad 1 \leq i, j \leq N, \tag{2.16}$$

The quantity $\mu(\boldsymbol{A})$ measures the smallest angle between any two columns of $\boldsymbol{A}$ can be. If there are two columns $\boldsymbol{a}_i = c\boldsymbol{a}_j$ for some scalar c, one can say that two columns are coherent and $\mu(\boldsymbol{A})$ reaches its maximum of 1. On the other hand, it is shown $\mu(\boldsymbol{A})$ is lower-bounded by

$$\mu(\boldsymbol{A}) \geq \sqrt{\frac{N-M}{M(N-1)}}$$

in the case of CS where $M < N$ [118]. Generally, a matrix with low coherence is desirable in CS system, as with low coherence, it guarantees that each column $\boldsymbol{a}_i$ samples elements in signal s_i to measure the original signal $\boldsymbol{s}$ nearly linear-independently and element-wise uniform in power so that the measurements carry nearly the same and maximum 'amount of information' of the sensed signal. The following theorems are seen in literature to give guarantees of exact or stable recovery via (2.9) or (2.10).

Theorem 2.4.2. *For a given matrix $\boldsymbol{A}_{M\times N} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$ where $\boldsymbol{\Phi}$ is presented by the product of a row selection operator $\boldsymbol{P}_{M\times N}$ which uniformly at random selects M rows out of N and a square sampling basis $\tilde{\boldsymbol{\Phi}}_{N\times N}$ ($\boldsymbol{\Phi} = \boldsymbol{P}\tilde{\boldsymbol{\Phi}}$), and the linear system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{s}$ where $\boldsymbol{s}$ is*

arbitrary signal with at most k non-zeros entries. If

$$M \geq C \cdot \mu^2(\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}^{-1}) \cdot kN \log N \tag{2.17}$$

for some constant C independent of k and N, the solution to 1-norm optimization (2.9) is exact, i.e. $\boldsymbol{s}^o = \boldsymbol{s}$*, with overwhelming probability [117].*

Even if the value of the constant C cannot be determined hence the minimum M required is not explicitly available, Theorem 2.4.2 gives the order of the required number of measurements $\mathcal{O}(\mu^2(\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}^{-1}) \cdot kN \log N)$ and it implies that for a given dimension of the sparse recovery problem, i.e. N and k, the mutual coherence of the product of the sampling and sparsifying basis, i.e. $\mu(\tilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}^{-1})$ becomes a direct indicator for compressibility in the context of CS.

The noisy case has been also considered by Candes and Plan, and a theorem of stable recovery related to mutual coherence is provided:

Theorem 2.4.3. *Consider a sensing matrix* $\boldsymbol{A}_{M\times N}$*,* $(M < N)$ *and the system* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{b}$*.* $\boldsymbol{s}$ *is a signal having k non-zero entries whose locations are drawn uniformly at random and values are with amplitudes* $|s_i| > (6 + \sqrt{2})\sigma\sqrt{2\log(M)}$ *and random signs.* $\boldsymbol{b}$ *denotes Gaussian noise with covariance of* σ^2*. If*

$$\mu(\boldsymbol{A}) < C/\log M, \tag{2.18}$$

and

$$k < \frac{CM}{||\boldsymbol{A}||_2^2 \log M}, \tag{2.19}$$

for some constant C independent of k and N, the solution to (2.10) is stable with overwhelming probability [119].

Then a more general RIPless guarantee of stable recovery is provided as follows.

Theorem 2.4.4. *Let* $\boldsymbol{s}$ *be an arbitrary fixed vector. Pick an arbitrary scalar* $\beta > 0$*.*

Then with the probability of at least $1 - 6/N - 6\exp(-\beta)$*, the solution to (2.12) with* $\lambda = 10\sqrt{\log N}\sigma$ *obeys*

$$||\boldsymbol{s} - \boldsymbol{s}^o||_2 \leq \min_{1\leq\kappa\leq k} C(1+\alpha)\left[\frac{\sigma_{[\kappa]}(\boldsymbol{s})}{\sqrt{\kappa}} + \sigma\sqrt{\frac{\kappa\log N}{M}}\right] \tag{2.20}$$

and

$$||\boldsymbol{s} - \boldsymbol{s}^o||_1 \leq \min_{1\leq\kappa\leq k} C(1+\alpha)\left[\sigma_{[\kappa]}(\boldsymbol{s}) + \sigma\kappa\sqrt{\frac{\log N}{M}}\right] \tag{2.21}$$

where the covariance of $\boldsymbol{b}$ *is* σ*, provided that* $M \geq C(1+\beta)\mu(\boldsymbol{A})k\log N$*.* C *is a constant and* α *is upperbounded by* $\log^{3/2} N$ *[116].*

Mutual coherence of a given matrix can be easily computed with polynomial complexity of N [107], with its value being the maximum amplitude of the off-diagonal entries of the Gram matrix $\boldsymbol{A}^H\boldsymbol{A}$, after each of $\boldsymbol{A}$'s columns is normalized by its 2-norm. Therefore it is more attractive for conveniently and quantitatively evaluating the quality of the sensing matrix. However, one major drawback of both RIPless coherence-based and RIP guarantees lies in the unknown constant, specifically C_1, C_2, C_3 and C_4 in Theorem 2.4.1 and the C in Theorem 2.4.2 and 2.4.3, which is significant in practically revealing the required number of measurements. However, they collectively tell the direction to find a matrix with better coherence to improve the compressibility and recovery performance.

2.4.3 Remarks on Sensing Matrices

It is reported that some random matrices, for example, the Gaussian matrix, and the Bernoulli matrix, satisfy the RIP by a large margin with at least high probability [120], and can also achieve nearly the lowest possible mutual coherence and hence provide the almost minimum required number of compressed measurements. Besides, these random matrices are universal in the sense that they are incoherent with all kinds of deterministic sparsifying matrices, leading to low mutual coherence of the sensing matrix $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$. For instance, with each entry in the sampling matrix $\boldsymbol{\Phi}$ are independent and identically distributed Gaussian random values, the entries of sensing matrix $\boldsymbol{A}$ also conform to

i.i.d. Gaussian distribution, regardless of the choice of $\boldsymbol{\Psi}$. It is stated [121] that if

$$M \geq \mathcal{O}(k \log (N/k)) \tag{2.22}$$

the sensing matrix $\boldsymbol{A}$ satisfies RIP (with $\delta_{2k} < \sqrt{2} - 1$) with a probability of failure decreasing to 0 exponentially fast with increasing M.

Despite the satisfactory performance in CS, these matrices with full random entries have major concerns in practical applications as it requires huge memory buffering for generating and storing these random entries and high complexity in the process of sampling due to their completely unstructured nature. Also, it should be pointed out that the structure of the sensing matrix is practically dictated by the hardware implementation of the sub-Nyquist sampler. With a particular practical sampler design, a structured sensing matrix should be considered, which are detailed with sub-Nyquist sampler architectures in Section 2.5. Similarly, it is also stated that for some structured random matrices, which are easier to implement or naturally exist in applications, for example randomly selected rows of a Fourier matrix, namely a partial Fourier matrix, also satisfy the RIP with failing probability decreasing to 0 exponentially fast with increasing M, with the best-known requirements on M being typically [122]

$$M \geq \mathcal{O}(k \cdot \text{polynomial}(\log (N/k))). \tag{2.23}$$

Apart from the above-mentioned recovery theories of CS, it is worth noting in particular that there exists a sufficient condition on the explicit relation between the sparsity and the required number of measurements to ensure the uniqueness of the sparse solution.

Theorem 2.4.5. *Given an arbitrary k-sparse signal* $\boldsymbol{s}$ *and compressed measurements* $\boldsymbol{y}$ *satisfying (2.7), the solution to (2.9) is uniquely* $\boldsymbol{s}^o = \boldsymbol{s}$ *if*

$$k \leq \frac{spark(\boldsymbol{A})}{2}, \tag{2.24}$$

where

$$spark(\boldsymbol{A}) := \arg\min_{\boldsymbol{d}\neq\boldsymbol{0}} \|\boldsymbol{d}\|_0 \ \ s.t. \ \ \boldsymbol{A}\boldsymbol{d} = \boldsymbol{0}$$

denotes the Kruskal rank of $\boldsymbol{A}$*.[123]*

The proof is straightforward through the definition of the Kruskal rank. This theorem, compared with those based on RIP and RIPless, has a much simpler form and the relationship between sparsity and the sensing matrix is explicit without some constant ambiguous in value. It implies that it is safe to design a sensing matrix with its Kruskal rank equal to or greater than twice the sparsity of the original signal to ensure a unique sparse solution. Although the theorem above, as a sufficient condition, gives rather conservative constraints on the minimum required number of measurements and maximum allowed sparsity, it is widely adopted in the discussion and practice of CS-based sub-Nyquist sampling system [61, 65, 80, 124].

It is then of interest to evaluate the Kruskal rank of the 'fat' sensing matrix $\boldsymbol{A}_{M\times N}$ where $M < N$, and find a $\boldsymbol{A}$ with full Kruskal rank, i.e. $\text{spark}(\boldsymbol{A}) = M$, namely a universal sampling pattern [109]. However, this is not a trivial task compared to its rank and mutual coherence, since calculating a Kruskal rank of a matrix requires a combinatorial search over all possible subsets of columns from $\boldsymbol{A}$, thus leading to exponential complexity with N [107]. Some useful conclusions on finding the universal sampling pattern $\boldsymbol{A}$ have been built for the particular structured case where $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}$, $\boldsymbol{\Phi}$ is a row selector and $\boldsymbol{\Psi}$ is an orthogonal basis: a prime number N dictates every pattern universal [125], and a $\boldsymbol{\Psi}$ drawing rows at uniformly random leads to universal $\boldsymbol{A}$ with high probability [126].

It is intriguing to see that the mutual coherence can be linked to the Kruskal rank via the following inequality [127]

$$\text{spark}(\boldsymbol{A}) \geq 1 + \frac{1}{\mu(\boldsymbol{A})}. \tag{2.25}$$

It has been proved [107] that by applying a more stringent constraint on the sparsity k using (2.25), a theorem parallel in the form of Theorem 2.4.5 also holds, and moreover, it can be extended to ensure exact reconstruction by 1-norm minimization.

Theorem 2.4.6. *Given an arbitrary k-sparse signal* $\boldsymbol{s}$ *and compressed measurements* $\boldsymbol{y}$ *satisfying (2.7), if*

$$k \leq \frac{1}{2} \cdot \left(1 + \frac{1}{\mu(\boldsymbol{A})}\right), \tag{2.26}$$

the solution to (2.9) is uniquely $\boldsymbol{s}^o = \boldsymbol{s}$*, which is also the unique solution to (2.8).*

This theorem maintains the simplicity in form as in Theorem 2.4.5, and more strikingly, contains only mutual coherence which is trivial to compute and preserves the uniqueness of solution for convex relaxation.

2.4.4 Recovery Algorithms

One crucial problem in the context of CS is the recovery of the sparse unknown signal from an underdetermined linear system which is possibly impaired by additive noise. Specifically, the task is to obtain a signal $\boldsymbol{s}^o$, which equals to or approximates (within a certain distance) the original sparse signal $\boldsymbol{s}$ that satisfies the noiseless or noisy CS model. Several typical categories of recovery algorithms have been proposed in the literature and are reviewed in the following.

2.4.4.1 1-norm Minimization

The classic approaches to find the sparse solution in CS are to directly solve the 1-norm minimization problems in Section 2.4.1. These approaches apply convex optimization techniques which have been also adopted in wide areas of application. There is no error on the convergence of these convex approaches as a result of there being only one local minimum, however, it may suffer from construction error due to the approximation to the l_0-norm in the objective. Performance bounds in CS are widely discussed based on 1-norm minimization problems as shown in Section 2.4.2, stating the conditions under which the 1-norm minimization are the exact or stable recovery of the original signal, i.e.

no construction error. The specific optimization techniques to various forms of 1-norm minimization objective are outlined as follows.

The unconstrained problem (2.12) Compared with all other forms of 1-norm minimization problems in Section 2.4.1, the optimization problem (2.12) is convex and an unconstrained problem hence it is straightforward to solve. Looking at the objective of (2.12), one can see it is the sum of a non-differentiable convex function $g(\boldsymbol{s}) := ||\boldsymbol{s}||_1$ and differentiable convex $r(\boldsymbol{s}) := \frac{1}{2\lambda}||\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}||_2^2$. For convenience, the problem (2.12) is rewritten as

$$\arg\min_{\boldsymbol{s}} g(\boldsymbol{s}) + r(\boldsymbol{s}). \tag{2.27}$$

The general idea for optimising of an unconstrained convex problem is to find the value of the variable to be optimised for which the differential equals zero or the sub-differential contains zero, for differentiable and non-differentiable objective respectively [128]. In the case of (2.27), specifically, such $\boldsymbol{s}$ is desired

$$\mathbf{0} \in \partial(g + r)(\boldsymbol{s}), \tag{2.28}$$

where $\partial(\cdot)$ denotes the sub-differential and $\mathbf{0} \in \partial(\cdot)(\cdot)$ denotes $\mathbf{0}$ is a member of the sub-differential set. Note that for differentiable function, the sub-differential is single-valued. From (2.28), the sub-differential operator can be split as such

$$\begin{aligned}
\mathbf{0} \in \partial(g(\boldsymbol{s}) + r(\boldsymbol{s})) &\iff \mathbf{0} \in \partial(g(\boldsymbol{s})) + \partial(r(\boldsymbol{s})) \\
&\iff \mathbf{0} \in (\boldsymbol{s} + \tau\partial(g(\boldsymbol{s}))) - (\boldsymbol{s} - \tau\partial(r(\boldsymbol{s}))) \\
&\iff (I - \tau\partial r)\boldsymbol{s} \in (I + \tau\partial f)\boldsymbol{s} \\
&\iff \boldsymbol{s} = (I + \tau\partial f)^{-1}(I - \tau\partial r)\boldsymbol{s},
\end{aligned} \tag{2.29}$$

where $\tau > 0$ is an arbitrary constant. From the last line in (2.29), such $\boldsymbol{s}$ satisfying this equation is called a fixed point, which could be found by the intuitive iterative algorithm

$$\boldsymbol{s}^{(i+1)} \leftarrow (I + \tau\partial f)^{-1}(I - \tau\partial r)\boldsymbol{s}^{(i)}. \tag{2.30}$$

The superscript $\cdot^{(i+1)}$ and $\cdot^{(i)}$ denote two consecutive numbers of iteration thereafter. This particular form of fixed point iteration is named forward-backward operator splitting, where $(I+\tau\partial f)^{-1}$ and $(I-\tau\partial r)$ are called forward and backward operator respectively. Denote intermediate variable $\boldsymbol{t}^{(i)} := (I-\tau\partial r)\boldsymbol{s}^{(i)}$. Specifically, for the forward operator on the variable of $\boldsymbol{t}^{(i)}$, one then has $\boldsymbol{s}^{(i+1)} = (I+\tau\partial f)^{-1}\boldsymbol{t}^{(i)}$ and $\boldsymbol{t}^{(i)} \in (I+\tau\partial f)\boldsymbol{s}^{(i+1)}$, which translates to

$$\mathbf{0} \in \boldsymbol{s}^{(i+1)} + \tau\partial g(\boldsymbol{s}^{(i+1)}) + \boldsymbol{t}^{(i)} \iff \mathbf{0} \in \partial\left(\tau g(\boldsymbol{s}^{(i+1)}) + \frac{1}{2}||\boldsymbol{s}^{(i+1)} - \boldsymbol{t}^{(i)}||_2^2\right), \tag{2.31}$$

which leads to the conclusion that $\boldsymbol{s}^{(i+1)}$ is the minimizer of the constructed function $\tau g(\boldsymbol{s}^{(i+1)}) + \frac{1}{2}||\boldsymbol{s}^{(i+1)} - \boldsymbol{t}^{(i)}||_2^2$, writing

$$\boldsymbol{s}^{(i+1)} = \arg\min_{\boldsymbol{s}} \tau g(\boldsymbol{s}) + \frac{1}{2}||\boldsymbol{s} - \boldsymbol{t}^{(i)}||_2^2. \tag{2.32}$$

Consider the specification $g(\boldsymbol{s}) = ||\boldsymbol{s}||_1$, and the problem (2.32) has a closed-form solution

$$\boldsymbol{s}^{(i+1)} = \text{shrink}(\boldsymbol{t}^{(i)}, \tau^{-1}), \tag{2.33}$$

where the shrinkage operation (also known as soft-thresholding) [129] on the element of $\boldsymbol{t}^{(i)} = \left[t_1^{(i)}, t_2^{(i)}, \cdots, t_N^{(i)}\right]$ is defined as

$$\text{shrink}(t_n^{(i)}, \tau^{-1}) := \begin{cases} \left|t_n^{(i)}\right| - \tau^{-1}, & t_n^{(i)} > \tau^{-1}, \\ 0, & -\tau^{-1} \leq t_n^{(i)} \leq \tau^{-1}, \\ \left|t_n^{(i)}\right| + \tau^{-1}, & t_n^{(i)} < -\tau^{-1}, \end{cases}$$

for $1 \leq n \leq N$, and for the vector

$$\text{shrink}(\boldsymbol{t}^{(i)}, \tau^{-1}) := \left[\text{shrink}(t_1^{(i)}, \tau^{-1}), \text{shrink}(t_2^{(i)}, \tau^{-1}), \cdots, \text{shrink}(t_N^{(i)}, \tau^{-1})\right].$$

Moreover, for the specific case $r(\boldsymbol{s}) = \frac{1}{2\lambda}||\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}||_2^2$, the solution to the backward step $\boldsymbol{t}^{(i)} := (I - \tau\partial r)\boldsymbol{s}^{(i)}$ is directly found by the closed-form derivative

$$\boldsymbol{t}^{(i)} = \boldsymbol{s}^{(i)} - \lambda^{-1}\boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y}). \tag{2.34}$$

Combining steps (2.33) and (2.34), one arrives at the fixed point iteration algorithm for the problem (2.12)

$$\boldsymbol{s}^{(i+1)} = \text{shrink}(\boldsymbol{s}^{(i)} - \lambda^{-1}\boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y}), \tau^{-1}), \tag{2.35}$$

where $\tau > 0$ introduced in (2.35) has the meaning of the step size of iterations.

The convergence properties of the routine (2.35) have been studied in [130], where the global convergence result from an arbitrary initial point under mild conditions is proved and the convergence rate is linear with $\tau \cdot \lambda$. As the step size τ is usually a small positive value to ensure convergence, the solver (2.35) can be very slow with small λ. Practical fixed point solvers with various continuation strategies have appeared in literature, and most notably being the fixed-point continuation (FPC) algorithm [130], where, to accelerate the convergence rate, starting from a large λ, a decreasing series of λ can be used over the iterations until it finally reaches the target λ.

In some literature, the fixed-point algorithm (2.35) is also referred to as a prox-linear algorithm [131], as the exactly same routine can be derived from the prox-linear techniques, which is another useful tool for solving convex problems with non-differentiable objectives. The idea of the prox-linear algorithm generally calls for the linearization (the first term of Taylor expansion) of the data fidelity term and an additional proximal term. For the problem (2.27), the prox-linear iteration writes

$$\begin{aligned}\boldsymbol{s}^{(i+1)} &\leftarrow \arg\min_{\boldsymbol{s}} g(\boldsymbol{s}) + r(\boldsymbol{s}^{(i)}) + \left(\nabla r(\boldsymbol{s}^{(i)})\right)^H \left(\boldsymbol{s} - \boldsymbol{s}^{(i)}\right) + \frac{\tau}{2}||\boldsymbol{s} - \boldsymbol{s}^{(i)}||_2^2 \\ &= \arg\min_{\boldsymbol{b}} g(\boldsymbol{s}) + \frac{\tau}{2}||\boldsymbol{s} - \left(\boldsymbol{s}^{(i)} - \tau^{-1}\nabla r(\boldsymbol{s}^{(i)})\right)||_2^2.\end{aligned} \tag{2.36}$$

After the insertion of $r(\boldsymbol{s}) = \frac{1}{2\lambda}||\boldsymbol{A}\boldsymbol{s}-\boldsymbol{y}||_2^2$ and $g(\boldsymbol{s}) = ||\boldsymbol{s}||_1$ back to (2.36), it immediately yields (2.35).

The constrained problem (2.9) One would argue that a very small λ, i.e. a very large penalization term could be used for solving (2.12) in order to approximate the noiseless recovery problem (2.9). However, the convergence of fixed-linear algorithm can be very slow with small λ in practice. It is also noted that even with a very large λ the solution to (2.12) and (2.9) are never consistent. Besides, some points in the sequence towards the optimization (2.9), though not solving (2.12) nor (2.9) in the noisy scenario, have 'better quality' in terms of sparsity and data fitness than the solutions of (2.9) [55]. Therefore, it is of great interest to solve (2.9).

The problem (2.9) is a typical convex optimization problem with equality constraint. A generic tool to convert this constrained optimization problem to an equivalent unconstrained problem is by finding the Lagrange dual through its Lagrangian function. Hereby the derivation of the algorithm solving (2.9) is briefly illustrated. From (2.9), one can write the Lagrangian as following

$$\mathcal{L}(\boldsymbol{s}, \boldsymbol{t}) = ||\boldsymbol{s}||_1 + \boldsymbol{t}^H(\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}), \tag{2.37}$$

where $\boldsymbol{t}_{M\times 1}$ denotes the Lagrange multiplier. The Karush-Kuhn-Tucker condition for (2.37), which is the necessary condition the optimal of the argument $\boldsymbol{s}^o$ and the Lagrange multiplier $\boldsymbol{t}^o$, writes

$$\begin{aligned} \nabla_{\boldsymbol{s}}\mathcal{L}(\boldsymbol{s}^o, \boldsymbol{t}^o) &= \boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{s}^o - \boldsymbol{y} &= \boldsymbol{0}. \end{aligned} \tag{2.38}$$

However, since the Lagrange itself is non-differentiable, the optimal is not directly available from the first equation in (2.38).

An alternative approach to solve constraint optimization problems is the quadratic

penalty method [132, 133], which proposes to add penalization term proportional to the violation of the constraint. The quadratic penalty function writes

$$\mathcal{Q}(\boldsymbol{s}, \mu) = ||\boldsymbol{s}||_1 + \frac{\mu}{2}||\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}||_2^2. \tag{2.39}$$

It is noted that (2.39) is exactly the objective of the unconstrained l_1 problem (2.12), where the constants follow $\mu := \lambda^{-1}$. The quadratic penalty method proposes to use an iterative routine. A series of increasing penalty constants $0 < \mu_1 < \cdots < \mu_i < \mu_{i+1} < \cdots$ are used in these iterations and at the ith iteration it approximately solves

$$\boldsymbol{s}^{(i+1)} \approx \arg\min_{\boldsymbol{s}} \mathcal{Q}(\boldsymbol{s}, \mu_i) = \arg\min_{\boldsymbol{s}} ||\boldsymbol{s}||_1 + \frac{\mu_i}{2}||\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}||_2^2, \tag{2.40}$$

where the approximation indicates that $\boldsymbol{s}^{(i+1)}$ should be found by some of the iteration steps towards the convex objective in (2.40), till $||\nabla_{\boldsymbol{s}} \mathcal{Q}(\boldsymbol{s}, \mu_i)|| < \delta_k$ is met. The steps for solving (2.40) can be the fixed-point iterations as detailed in Section 2.4.4.1, Paragraph 'The Unconstrained Problem (2.12)'.

The idea of taking $\mu_i \to +\infty$ can lead to an optimal that asymptotically converges to satisfy the KKT conditions (2.38). However, the indefinitely large value of μ_i is prone to resulting in an ill-conditioned Hessian $\nabla^2_{\boldsymbol{ss}} \mathcal{Q}(\boldsymbol{s}, \mu_i)$, making the iterative steps less effective and prone to numerical errors [132, 133]. Moreover, in practice it could be another problem that how to choose a properly large μ_i.

To mitigate the problem of ill-conditioning and choosing a large μ_i, the augmented Lagrange method is used, which is closely related to the original form of quadratic penalty method [132, 134]. In analogy to the quadratic penalty function, the augmented Lagrange function preserves the quadratic penalization term and also includes the Lagrange multiplier, which writes

$$\mathcal{L}_{\mathcal{A}}(\boldsymbol{s}, \boldsymbol{t}, \mu) := \mathcal{Q}(\boldsymbol{s}, \mu_i) + \boldsymbol{t}^H(\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}) = ||\boldsymbol{s}||_1 + \boldsymbol{t}^H(\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}) + \frac{\mu}{2}||\boldsymbol{A}\boldsymbol{s} - \boldsymbol{y}||_2^2. \tag{2.41}$$

By iteratively optimizing (2.41) approximately, specifically,

$$\boldsymbol{s}^{(i)} \approx \arg\min_{\boldsymbol{s}} \mathcal{L}_{\mathcal{A}}(\boldsymbol{s}, \boldsymbol{t}^{(i)}, \mu_i), \tag{2.42}$$

the argument at the ith iteration satisfies

$$\nabla_{\boldsymbol{s}} \mathcal{L}_{\mathcal{A}}(\boldsymbol{s}^{(i)}, \boldsymbol{t}^{(i)}, \mu_i) = \nabla_{\boldsymbol{s}}(||\boldsymbol{s}^{(i)}||_1) + \boldsymbol{A}^H \left[\mu_i(\boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y}) - \boldsymbol{t}^{(i)}\right] \approx 0. \tag{2.43}$$

In order to make the KKT satisfies approximately, recall the first formula in (2.38), which is related to (2.43), and one immediately has

$$\begin{aligned} \boldsymbol{t}^o &\approx \boldsymbol{t}^{(i)} - \mu_i(\boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y}) \\ \boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y} &\approx \mu_i^{-1}(\boldsymbol{t}^{(i)} - \boldsymbol{t}^o), \end{aligned} \tag{2.44}$$

where the violation to the second KKT condition (2.38) can be asymptotically diminished by the convergence of the Lagrange multiplier $\boldsymbol{t}^{(i)}$, thus not requiring an indefinitely large μ_i. This effectively prevents ill-conditioning of the step (2.44). In light of (2.44), the update for the Lagrange multiplier is given by intuition, aiming to find an estimate of the optimum Lagrange multiplier $\boldsymbol{t}^o$,

$$\boldsymbol{t}^{(i+1)} \leftarrow \boldsymbol{t}^{(i)} - \mu_i(\boldsymbol{A}\boldsymbol{s}^{(i)} - \boldsymbol{y}). \tag{2.45}$$

Iteration steps (2.43) and (2.45) compose the augmented Lagrangian algorithm for constrained problem (2.10), where μ_i does not need to increase indefinitely with i to ensure convergence and can stay at a much smaller value with comparison to the quadratic penalty method. For the step (2.43), fixed-point techniques for the unconstrained l_1 problem (2.12) can be applied. There exists a threshold of μ_i above which the solution of the augmented Lagrangian algorithm asymptotically converges to the optimizer of the problem (2.10), and a larger μ_i accelerates the convergence [132, 134]. In practice, μ_i should be empirically chosen to be above the threshold and compromise between convergence rate and the chance of ill-conditioning.

It is noted that the iterative algorithm of the augmented Lagrange method for the problem (2.10) coincide with that of the Bregman method [135], although they are derived using separate theories of convex optimization techniques.

Remark In the above paragraphs, only the basic ideas and techniques of the l_1 optimization problem are introduced. The topic of optimization techniques of l_1-related problems has been widely discussed not only in the applications of sparse signal recovery but also in statistics, regression, etc. For example, the objective form (2.11) is referred to as the LASSO problem in the context of regression, which has a classic solver namely spectral projected gradient (SPG) method [85]. Another commonly-used strategy for solving LASSO is via homotopy/LARS method [136, 137], which proposes to find all the solutions to the unconstrained problem (2.12) for a series of decreasing parameters λ using an active-set continuation strategy, and eventually finds the solution to the equivalent LASSO problem (2.11) satisfying the constraint with the parameter τ. The same strategy also applies for solving the inequality constraint problem (2.12), as the problems (2.10), (2.11) and (2.12) share the same solution for some fixed parameters σ, τ and λ respectively [138]. Another continuation strategy for (2.12) is also seen in literature, where a series of LASSO problems (2.11) with increasing τ's are proposed to be solved, each by spectral projected gradient (SPG) algorithm [138]. Moreover, the alternating direction method of multiplier (ADMM) method, inheriting the idea of augmented Lagrange method, is also a commonly used framework for l_1 problems such as (2.10), (2.11) and (2.12) [139, 140]. In ADMM, the objective is proposed to be split into the linear combination of two functions of different variables and so as the constraint. Then, the optimization of the augmented Lagrangian is proposed to be performed with regards to each variable, where one of the two subproblems could be solved approximately to accelerate the algorithm.

2.4.4.2 Non-convex minimization

Another approach to approximate the NP-hard l_0 norm minimization is to relax the objective to the l_v norm ($0 < v < 1$) to form a non-convex objective. Specifically, define $l_v := \left(\sum_{n=1}^{N} |s_n|^v\right)^{\frac{1}{v}}$, and the objective writes

$$\boldsymbol{s}^o = \arg\min_{\boldsymbol{s}} ||\boldsymbol{s}||_v^v + \frac{1}{2\lambda}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2. \tag{2.46}$$

Compared with the l_1 norm convex relaxation, the global optimum of the l_v objective can achieve the exact reconstruction with substantially fewer measurements, due to the better approximation to the original l_0 objective [141]. However, due to the non-convex nature of the objective, the l_v norm recovery are prone to undesired convergence to some local optimum and has to be solved by iteratively reweighted least square (IRLS) algorithms (also termed as the FOcal Underdetermined System Solver, FOCUSS) [142]. Each iteration updates the weights on the elements of the signal and solves a least-square problem by classic convex optimization solver. Specifically, the iterations have the following form

$$\begin{aligned} \boldsymbol{s}^{(i)} &\leftarrow \arg\min_{\boldsymbol{s}} \sum_{n=1}^{N} w_n^{(i)} s_n s_n^* + \frac{1}{2\lambda}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2, \\ w_n^{(i)} &\leftarrow \left(\left(s_n^{i-1}\right)^2 + \rho\right)^{\frac{v}{2}-1}, \ 1 \le n \le N, \end{aligned} \tag{2.47}$$

where ρ is a small positive regularization term to prohibit infinite weights. The first least-square subroutine has the analytic solver

$$\boldsymbol{s}^{(i)} \leftarrow \boldsymbol{W}^{(i)}\boldsymbol{A}^H \left(\boldsymbol{A}\boldsymbol{W}^{(i)}\boldsymbol{A}^H + \lambda\mathbf{I}\right)\boldsymbol{y}, \tag{2.48}$$

where $\boldsymbol{W}^{(i)}$ is a $N \times N$ diagonal matrix with $w_n^{(i)}$ being the nth diagonal element. Variants of the primitive IRLS adopt continuation strategies on the update of regularization parameters ρ and λ to ensure and accelerate the convergence [143, 144]. Despite the possibility of convergence to local mimima, the overall recovery performance of l_v norm

optimization can still be superior than convex relaxation with less stringent requirements on the compression ratio and signal sparsity.

2.4.4.3 Greedy Algorithms

The basic idea behind the greedy algorithms is that in every iteration the algorithm detects one index or a few of indices of the sparse signal $\boldsymbol{s}$ that are most likely to be non-zero by correlating the measurement vector with every column of the sensing matrix. The basic and most commonly adopted greedy algorithm is orthogonal matching pursuit (OMP) [145, 146], where the support of $\boldsymbol{s}$ is sequentially detected in each iteration. OMP algorithm implements the heuristic by incrementally pursuing a new support candidate and finding the optimal solution that describe the detected supports. With empty initial support $\Omega^{(0)} = \varnothing$ and $\boldsymbol{s}^{(0)} = \boldsymbol{0}$, the ith iteration updates the residual $\boldsymbol{e}^{(i)}$, the detected support $\Omega^{(i)}$ and the recovered signal $\boldsymbol{s}^{(i)}$ by the following procedures:

$$\boldsymbol{e}^{(i)} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}^{(i-1)} \quad \text{(OMP, residual)}, \tag{2.49}$$

$$\Omega^{(i)} = \Omega^{(i-1)} \cup \arg\max_{\theta} \left\{ \left|\left| \boldsymbol{a}_\theta^H \boldsymbol{e}^{(i)} \right|\right|_2 : \theta \notin \Omega^{(i-1)} \right\} \quad \text{(OMP, correlation, support update)} \tag{2.50}$$

$$\left(\boldsymbol{s}^{(i)}\right)_{\Omega^{(i)}} = (\boldsymbol{A})_{\Omega^{(i)}}{}^{\dagger} \boldsymbol{y}, \left(\boldsymbol{s}^{(i)}\right)_{\overline{\Omega^{(i)}}} = \boldsymbol{0} \quad \text{(OMP, least-squares)} \tag{2.51}$$

where subscription $(\cdot)_\Omega$ denotes a sub-vector or sub-matrix constructed by columns with indices in set Ω, and $(\cdot)^\dagger$ represents pseudo inverse. $\boldsymbol{a}_\theta$ denotes the θth column of the sensing matrix $\boldsymbol{A}$. Note that OMP, along with most of other greedy algorithms, requires the sparsity level k as the prior input to the algorithm to determine the maximum number of iterations. Before reaching the maximum number of iteration, the algorithm can also be terminated by a smaller enough (by predefined threshold) residual.

Besides the OMP algorithm, other greedy algorithm featuring low complexity and support detection can be also applied to perform CS recovery, namely Compressive Sampling Matching Pursuit (CoSaMP) [147], Subspace Pursuit (SP) [148], and Hard

Thresholding Pursuit (HTP) [149], and all these algorithms also require estimated sparsity order as the input. The difference between OMP and these three algorithms is that the OMP cumulatively detects the support in the desired signal, while CoSaMP, SP, and HTP update a certain number (equal to the specified sparsity order) of supports batch by batch. A notable variant of OMP - stagewise-OMP (StOMP) [150] is proposed in the enlightenment of dealing with noisy residuals in signals for greedy pursuit algorithms. The main difference from the original OMP is that StOMP incrementally selects multiple non-zero indices into the support using a soft thresholding strategy while the original OMP incrementally selects the most prominent one in each iteration. The main observation and proved argument in [150] is that the residual correlation in each OMP step contains a small number of significant non-zeros in addition to a vector disturbed by Gaussian-like noise in each entry. The threshold in each StOMP's correlation step is related to the power of the residual, and in noisy cases, also the noise variance. The major significance of StOMP is that it does not require the sparsity input and that it only runs a fixed number of iterations (e.g. 10 suggested in [150]) which can be much smaller than that of the original OMP. However, it requires the scale of the problem to be large, i.e. N and M are both large to approximately satisfy the Gaussianity assumption, where the typical values used for validation in [150] are $M = 256$ and $N = 1024$.

2.4.4.4 Sparse Bayesian Learning

The underlying assumption of the l_v optimization ($0 < v \leq 1$) in sparse recovery is that the deterministic form of the sensed signal, that is the signal $||\boldsymbol{s}||_0$ is small. However, originally proposed to be applied to support vector machine problem which has similar linear model and the need of encouraging sparsity, the SBL strategy has been proposed in sparse signal recovery problem in the perspective of probability [91]. Given the general CS model (2.7), SBL assumes a Gaussian prior model of the signal to be sensed, :

$$p(\boldsymbol{s};\boldsymbol{\gamma}) = \mathcal{N}(\boldsymbol{s}|\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma})) \tag{2.52}$$

where $\boldsymbol{\gamma} = [\gamma_1\ \gamma_2\ \cdots\ \gamma_N]$ is defined as a vector of prior hyper-parameters. It is noted that $\boldsymbol{s}$ is modelled as a statistical signal, where $\mathcal{N}(\boldsymbol{s}|\mathbf{0}, \mathrm{diag}(\boldsymbol{\gamma}))$ stands for the probability density function of Gaussian distribution of variables $\boldsymbol{s}$ with means $\mathbf{0}$ and covariance matrix $\mathrm{diag}(\boldsymbol{\gamma})$. Moreover, considering the Gaussian noise $\boldsymbol{b}$ with variance σ^2, the probability density function of measurements $\boldsymbol{y}$ conditioned on $\boldsymbol{s}$ writes

$$p(\boldsymbol{y}|\boldsymbol{s};\sigma) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{s}, \sigma^2\mathbf{I}). \tag{2.53}$$

SBL aims to maximize the marginal PDF (marginal likelihood) with σ and hyperparameters $\boldsymbol{\gamma}$

$$(\sigma_{\mathrm{ML}}, \boldsymbol{\gamma}_{\mathrm{ML}}) = \arg\max_{\sigma,\boldsymbol{\gamma}} p(\boldsymbol{y};\boldsymbol{\gamma},\sigma) = \arg\max_{\sigma,\boldsymbol{\gamma}} \int p(\boldsymbol{y}|\boldsymbol{s};\sigma)p(\boldsymbol{s};\boldsymbol{\gamma})\mathrm{d}\boldsymbol{s}. \tag{2.54}$$

The maximization of the integral in (2.54) is referred to as evidence maximization or type-II maximum likelihood [91, 94]. The integral term can be expressed by

$$p(\boldsymbol{y};\boldsymbol{\gamma},\sigma) = (2\pi)^{-(N/2)}|\boldsymbol{\Sigma}_y|^{-(1/2)}\exp\left\{-\frac{1}{2}\boldsymbol{y}^t\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}\right\}, \tag{2.55}$$

where $\boldsymbol{\Sigma}_y := \sigma^2\mathbf{I} + \boldsymbol{A}\cdot\mathrm{diag}(\boldsymbol{\sigma})\cdot\boldsymbol{A}^H$ [91]. Then the objective of SBL problem can be written in the equivalent form by taking the logarithm of the marginal PDF, i.e. the likelihood function

$$(\sigma_{\mathrm{ML}}, \boldsymbol{\gamma}_{\mathrm{ML}}) = \arg\max_{\sigma,\boldsymbol{\gamma}} \log\{p(\boldsymbol{y};\boldsymbol{\gamma},\sigma)\} = \arg\max_{\sigma,\boldsymbol{\gamma}} \log|\boldsymbol{\Sigma}_y| + \boldsymbol{y}^t\boldsymbol{\Sigma}_y^{-1}\boldsymbol{y}. \tag{2.56}$$

The objective in (2.56) is non-convex [94]. To find a local optimum for such non-convex likelihood objective with unobserved latent data s that has been integrated out as in (2.54), a common approach is to apply the Expectation-Maximization (EM) algorithm [151]. The basic idea behind the EM framework is to find a convex surrogate function approximating the original objective, and an optimum of the original objective can be found by iteratively optimizing the surrogate function until convergence. In this setting,

the surrogate function takes the expectation form of following

$$Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^{(i)}) = \mathrm{E}_{\boldsymbol{s}|\boldsymbol{y};\sigma,\boldsymbol{\gamma}^{(i)}} p(\boldsymbol{y}|\boldsymbol{s};\sigma) p(\boldsymbol{s};\boldsymbol{\gamma}), \tag{2.57}$$

where the superscript $(\cdot)^{(i)}$ represents the parameters in the ith iteration and $\mathrm{E}_{\boldsymbol{s}|\boldsymbol{y};\sigma,\boldsymbol{\gamma}^{(i)}}(\cdot)$ denotes the expectation of a random variable with the conditional probability distribution $f(\boldsymbol{s}|\boldsymbol{y};\sigma,\boldsymbol{\gamma}^{(i)})$. Using the Bayesian rule, the posterior probability term in $p(\boldsymbol{s}|\boldsymbol{y};\sigma,\boldsymbol{\gamma})$ is accounted by

$$p(\boldsymbol{s}|\boldsymbol{y};\sigma,\boldsymbol{\gamma}) = \frac{p(\boldsymbol{s};\boldsymbol{\gamma}) p(\boldsymbol{y}|\boldsymbol{s};\sigma)}{p(\boldsymbol{y};\boldsymbol{\gamma},\sigma)} = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\eta},\boldsymbol{\Sigma}_s), \tag{2.58}$$

where $\boldsymbol{\eta} := \sigma^{-2}\boldsymbol{\Sigma}_s \boldsymbol{A}^H \boldsymbol{y}$ and $\boldsymbol{\Sigma}_s := \left\{\sigma^{-2}\boldsymbol{A}^H\boldsymbol{A} + [\mathrm{diag}(\boldsymbol{\gamma})]^{-1}\right\}^{-1}$ [91]. Then, the maximization step calculates the updates for the parameters $\boldsymbol{\gamma}$ and σ by maximizing the surrogate function $Q(\boldsymbol{\gamma}, \boldsymbol{\gamma}^{(i)})$

$$\begin{aligned} \gamma_n^{(i+1)} &= [\boldsymbol{\Sigma}_s^{(i)}]_{n,n} + (\eta_n^{(i)})^2, \quad 1 \le n \le N \\ (\sigma^2)^{(i+1)} &= \frac{||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}||_2^2 + (\sigma^2)^{(i)} \sum_{j=1}^{N} \left[1 - [\boldsymbol{\Sigma}_s^{(i)}]_{j,j}/\gamma_j^{(i)}\right]}{M}, \quad 1 \le n \le N \end{aligned} \tag{2.59}$$

where $[\boldsymbol{\Sigma}_s^{(i)}]_{n,n}$ is the nth diagonal element of $[\boldsymbol{\Sigma}_s^{(i)}]$ calculated by definition by $\boldsymbol{\gamma}^{(i)}$ and $(\sigma^2)^{(i)}$, and $\eta_n^{(i)}$ is the nth element of $\boldsymbol{\gamma}^{(i)}$ calculated by definition by $\boldsymbol{\gamma}^{(i)}$ and $(\sigma^2)^{(i)}$.

Upon convergence, $(\sigma_{\mathrm{ML}}, \boldsymbol{\gamma}_{\mathrm{ML}})$ is revealed and $\boldsymbol{s}^o$ is then obtained as the expectation of the posterior distribution (2.58)

$$\boldsymbol{s}^o = \boldsymbol{\eta}_{\mathrm{ML}} = \left(\boldsymbol{A}^H\boldsymbol{A} + \sigma_{\mathrm{ML}}^2 \cdot \mathrm{diag}(\boldsymbol{\gamma}_{\mathrm{ML}})\right) \boldsymbol{A}^H \boldsymbol{y}. \tag{2.60}$$

The global optimum of the SBL's objective is proved to correspond to the sparsest solution, other than the convex and non-convex relaxation objective where the global optimum is not generally the sparsest solution [87], which means the recovery is more accurate when the global optimum is found, i.e. less structural error [94]. Moreover, compared to non-convex relaxation (l_v-norm), the SBL's objective has explicitly less local minima, hence less chance of occurring convergence error [94]. Numerical simulation in

general settings reveals superior recovery performance against both basis pursuit and FOCUSS, taking both types of error into account. The superiority is reportedly evident especially in the presence of highly-correlated columns of the sensing matrix $\boldsymbol{A}$, poor signal sparsity and low SNR [94].

2.5 Compressive Sensing Frameworks for Multiband Signals

In this section, three popular and practical sub-Nyquist samplers are shown, with the corresponding CS mathematical models and practicality concerns of them reviewed. At the end of this section, the implications of compressive sensing techniques on the receiver performance are presented.

2.5.1 Random Demodulator

The input signal of the random demodulator is assumed to be a multitone signal $s(t)$ that is frequency-sparse and bandlimited within $\left[-\frac{B}{2}\ (\text{Hz}), \frac{B}{2}\ (\text{Hz})\right]$ in the frequency domain. In a random demodulator system, the input signal is multiplied by a Nyquist-rate sequence from a pseudorandom chip generator. As a result, the original signal is modulated pseudorandomly, in other words, scrambled. Then, after an integration stage, the modulated signal is proposed to be sampled with sub-Nyquist rate and produces discrete-time samples $\boldsymbol{y} = [y_1, y_2, \cdots, y_M]$.

Note the pseudorandom sequence by the following waveform with a chip rate of $1/T_s = B$ (Hz) at Nyquist rate

$$p(t) = \sum_{i=0}^{M-1} \beta_i \text{rect}_{T_s}(t - i \cdot T_s), \quad 0 \leq t < NT_s, \tag{2.61}$$

where $\text{rect}_{T_s}(t - i \cdot T_s)$ is the rectangular waveform whose value is 1 if and only if $1 \leq t < T_s$, and β_i denotes the amplitude of the ith rectangular waveform.

Following the signal model and analysis of the sampling process provided by Tropp et. al. [71], the compressive sampling of a continuous-time multitone signal $s(t)$ can be effectively represented by a simple matrix form in accordance with the general CS model (2.7) using discrete time-domain and frequency-domain signal representations, noted by $\boldsymbol{s}$ and $\boldsymbol{x}$ respectively. The rationale of such 'continous-to-discrete' translation and relationships among $s(t)$, $\boldsymbol{s}$ and $\boldsymbol{x}$ are detailed in [71]. A detailed modeling of the random demodulator's sampling process with a continuous signal model is given in [73]. Using the discrete signal representations $\boldsymbol{s}$ and $\boldsymbol{x}$ proposed by Tropp et. al. [71], one can relate the sampling process of the random demodulator with a sensing matrix $\boldsymbol{A}$ of the CS system (2.7) in the following form

$$\boldsymbol{A} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{\Psi}^{-1}, \tag{2.62}$$

where the block matrix $\boldsymbol{H}_{M\times N}$ represents the integral process whose rth row has M/N consecutive elements of 1s from the $(rM/N+1)$th column to the $((r+1)M/N)$th column, for each integer $0 \leq r \leq M/N-1$. The effective sampling matrix $\boldsymbol{\Phi}$ as defined in the general CS model (2.7) is $\boldsymbol{\Phi} = \boldsymbol{H}\boldsymbol{D}$. The sparsifying matrix $\boldsymbol{\Psi}^{-1}$ here is the inverse discrete Fourier transform matrix (IDFT). $\boldsymbol{D} = \mathrm{diag}(\beta_1, \beta_2, \cdots, \beta_M)$ is the matrix corresponding to the periodic pseudorandom sequence. Rewrite (2.7) using (2.62) and one has

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{\Psi}^{-1}\boldsymbol{x} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{s} \tag{2.63}$$

where $\boldsymbol{y}$ is the measurement vector; $\boldsymbol{x}$ is the discrete frequency-domain representation of the multitone signal $s(t)$; and $\boldsymbol{s} = \boldsymbol{\Psi}^{-1}\boldsymbol{x}$ is the discrete time-domain representation of the multitone signal $s(t)$. With the sampling model (2.63), one can intuitively understand the sampling process in the time domain. Firstly, the multiplication of the diagonal matrix $\boldsymbol{D}$ and the time-domain signal $\boldsymbol{s}$ corresponds to the mixing of the input signal and pseudorandom waveform. Next, the block-diagonal matrix $\boldsymbol{H}$ performs the 'integral and dump' operation on the scrambled signal $\boldsymbol{D}\boldsymbol{s}$, with each row of $\boldsymbol{H}$ corresponding to one integral interval. The inner product of the nth ($n = 1, 2, \cdots, M$) row of $\boldsymbol{H}$ and the

scrambled signal $\boldsymbol{Ds}$ turns out to be the nth output sample y_n, i.e. the nth element in the measurement matrix $\boldsymbol{y}$. In addition to the 'mix, intergrate and dump' intepretation in the time domain, the sampling procedure can be viewed in frequency domain such that the original sparse signal is convolved by various pseudorandom frequency coefficients into multiple version of 'signature', effectively forming an underdetermined linear system [73].

From the idea of Fourier analysis, any continuous time-domain signal can be decomposed by an infinite number of monotone components with different (complex) weights. In some applications, a real-world signal can be approximated by the weighted sum of only a limited number of monotones, i.e. a multitone signal, which fits in the signal model and hence can be applied to the random demodulator. However, such approximation using the multitone model causes one underlying disadvantage of the random demodulator - some frequency-sparse analog signals require a great number, i.e. N, of monotone components to be well approximated and provide reasonable spectral resolution in the discrete model, which in other word requires a large dimension N of the sparse signal $\boldsymbol{s}$ and hence large CS recovery complexity [65].

Due to the causality of the 'integration and dump' process, the matrix $\boldsymbol{H}$ in the CS model (2.63) is restricted to the form of block diagonal matrix, which limits the

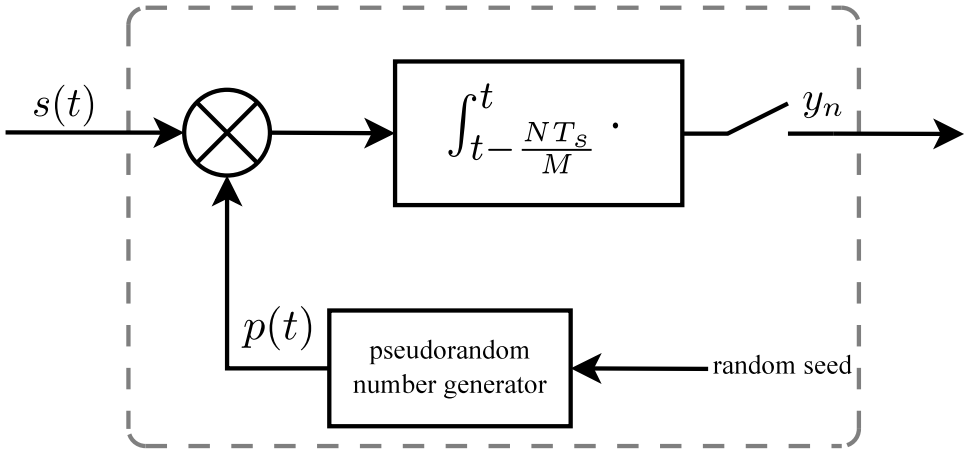


Figure 2.8: AIC architecture.

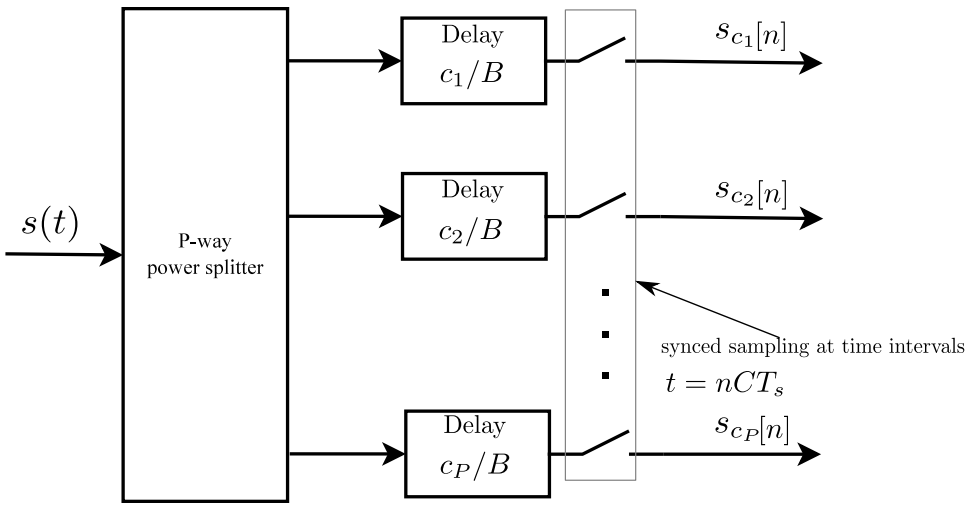


Figure 2.9: Multicoset sampler architecture.

performance of CS in comparison to a full random matrix. Some literature suggests multiple channels (i.e. duplicates of the outlined block) of the random demodulator to form more complex sensing matrices towards full Bernoulli [76, 78]. The properties of the sensing matrix $\boldsymbol{A} = \boldsymbol{HD\Psi}$ and the performance bounds in the context of CS of the random demodulator have been investigated in [71].

2.5.2 Multi-Coset Sampler

Multi-coset sampler, however, adopts another sampling architecture by using multiple branches of sub-Nyquist-rate ADCs to achieve non-uniform sampling [109], as shown in Fig. 2.9. Each of the P cosets has an ADC sampling at a fraction-L of the Nyquist frequency.

Consider the baseband frequency-sparse signal $s(t)$ bandlimited to the frequency range of $\left[-\frac{B}{2}\ (\text{Hz}), \frac{B}{2}\ (\text{Hz})\right]$, where there exist C channels with equal bandwidth of $\frac{B}{C}$. The sampling procedure of the p ADCs produces

$$s_{c_p}[n] = s\left(\frac{nC - c_p}{B}\right) \tag{2.64}$$

at the sub-Nyquist rate of B/C, for $n = \mathbb{Z}$ and $p = 1, 2, \cdots, P$. The unique delay for the

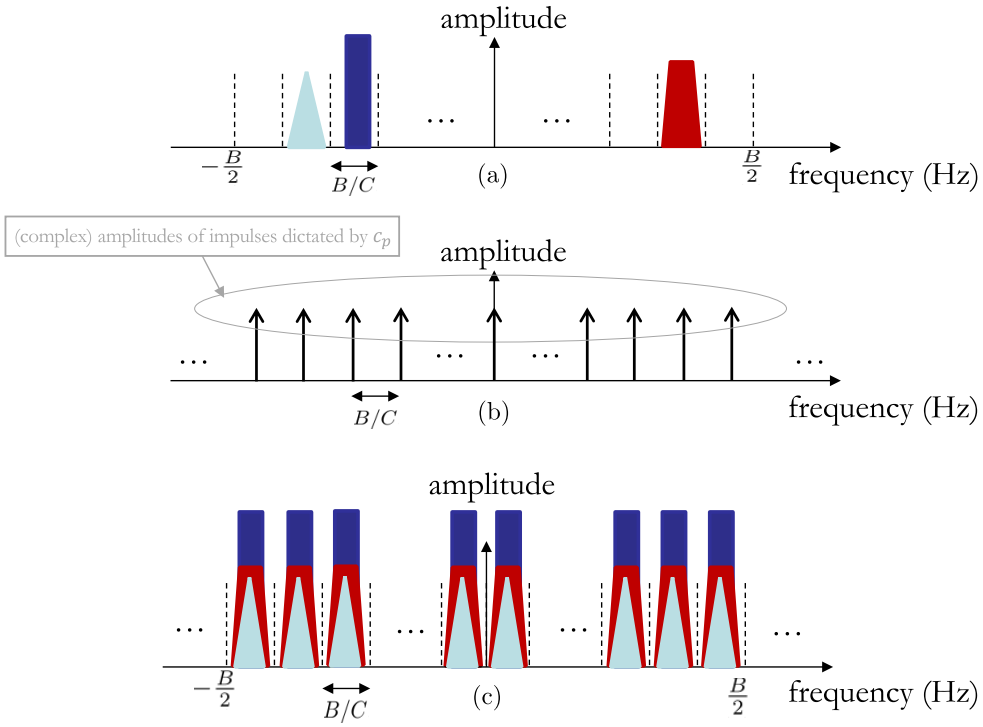


Figure 2.10: Frequency-domain illustration of the sampling process of the multicoset sampler. **(a)** The spectrum of the input multiband signal $s(t)$ with the channels sparsely occupied; **(b)** the Fourier transform of the time-shifted sampling pulse train of the pth channel; **(c)** the Fourier transform of the discrete samples $s_{c_p}[n]$'s, as the convolution of (a) and (b).

pth ADC is $\frac{c_p}{B}$, where $0 \leq c_p \leq C-1$ is an exclusively chosen integer, i.e. $c_p \neq c_q$ for $p \neq q$. The discrete samples $s_{c_p}[n]$'s are collected, for $n = \mathbb{Z}$ and $p = 1, 2, \cdots, P$. Intuitively, the sampling process can be interpreted in the frequency domain as illustrated in Fig. 2.10. The sub-Nyquist sampling of the delayed input signal can be equivalently regarded as that multiple identical versions of $s(t)$ are sampled (multiplied) by different time-shifted versions of a periodic impulse train. In the frequency domain, each time-domain impulse train can be represented by a series of impulses, whose (complex) amplitudes are determined by the delay value of the particular branch, as shown in Fig. 2.10 (b). The resultant discrete samples $s_{c_p}[n]$'s of each branch then have the frequency-domain representation of the multiband signal's spectrum convolved by these frequency-domain impulses, effectively creating a linear combination of the spectra from each channel, as illustrated in Fig. 2.10 (c). Without detailing the derivations, the CS model of such

processing architecture is hereby given directly

$$\boldsymbol{y}(f) = \boldsymbol{A}\boldsymbol{x}(f), f \in \left[0, \frac{B}{C}\right) \tag{2.65}$$

where $\boldsymbol{y}(f) = [\boldsymbol{Y}_{c_1}(f), \boldsymbol{Y}_{c_2}(f), \cdots, \boldsymbol{Y}_{c_P}(f)]^T$ is the measurement vector directly related to the frequency-domain representations of $s_{c_p}[n]$'s - $\boldsymbol{Y}_{c_p}(f) := \frac{C}{B} e^{j2\pi c_p \frac{f}{B}} \cdot X_{c_p}\left(e^{j2\pi f \frac{C}{B}}\right)$ and $X_{c_p}\left(e^{j2\pi f \frac{C}{B}}\right)$ is the discrete-time Fourier transform (DTFT) of $s_{c_p}[n]$'s. The vector $\boldsymbol{x}(f) := \left[X(f - \frac{B}{2}), X(f - \frac{B}{2} + \frac{B}{C}), \cdots, X(f + \frac{B}{2} - \frac{B}{C})\right]^T$ has its cth-element being the 'frequency-shifted' version of $s(t)$'s spectrum, $X(f)$, representing the spectrum of the cth channel. The entries in the matrix $\boldsymbol{A}_{P \times C}$ are expressed by $\{\boldsymbol{A}\}_{p,q} = \exp j2\pi c_p \left(\frac{1}{2} - \frac{q-1}{C}\right)$, for $1 \leq p \leq P$ and $1 \leq q \leq C$. Each row of $\boldsymbol{A}$ is related to the cth branch's delay c_p, whose values are reflected by the Fourier series of the time-shifted sampling impulse train. The inner product between each row of $\boldsymbol{A}$ and the shifted spectrum $\boldsymbol{x}(f)$ corresponds to the sampling process of each branch, more intuitively, the convolution between the input signal's spectrum and the Fourier series of the sampling impulse train, where the frequency-domain representation of $s_{c_p}[n]$'s is a linear combination of the spectrum from each channel. Readers are referred to Section 5.1 for detailed evidences of such mathematical model of the multicoset sampler.

The model (2.65) contains a parameter f whose range is a continuous period. The number of measurement vectors with every applicable f is infinite, so the model (2.65) is often referred to as an infinite measurement vector (IMV) model. In order to apply CS recovery algorithms, the IMV model can be simplified to an MMV model by selecting a group of discrete f's to provide a reasonable spectrum resolution or subspace analysis techniques [80]. For the MMV model, the recovery based on the MMV model is proposed to be efficiently achieved by simultaneous greedy algorithms, such as SOMP and AR-SOMP [16], as well as the simultaneous version of other sparse recovery techniques [152, 153], e.g. M-SBL [154] and M-FOCUSS [45]. Discussions are also present for how to design a series of delay values, i.e. c_p's to achieve a good sensing matrix, where it indicates that choosing these values exclusively and uniformly random among integers

$0 \leq c_p \leq C-1$ is a reasonable practice to construct a good sensing matrix $\boldsymbol{A}$ [65].

The architecture of the multicoset sampler has a simple structure and can be realized by general-purpose ADC devices. It can also form an MMV CS model, where the model dimension and the recovery complexity may be significantly lower than the SMV model of the random demodulator to achieve the same spectral resolution. However, it is addressed in [65] that there are practicality issues regarding the usage of sub-Nyquist-rate ADCs in the multicoset sampler. Specifically, apart from the sampling rate, ADC devices also have a practical limitation on front-end bandwidth, the frequency components exceeding which are lost, potentially causing inaccurate sampling results. In each branch of the multicoset sampler, the sampling rate is a fraction of the Nyquist-rate but the analog input signal $s(t)$ seen by the ADC has to have a bandwidth equal to the Nyquist rate of B Hz. ADCs with such a difference between sampling rate and front-end bandwidth (referred to as full-power bandwidth, FPBW) need particular attention in design and not commonly seen. In Fig. 2.11, a chart of wideband ADCs (sampling rate no less than 500 MHz) from Analog Devices product catalog is illustrated. It can be seen that the ratio of FPBW against maximum sampling rate is small (all less than 5), meaning that one has to choose ADCs with excessively high maximum sampling rate to satisfy the required FPBW, but the ADCs are actually running at a low sampling rate. This is effectively a waste of the hardware's capability and cost. Moreover, the delays of the sampling timing are at the resolution corresponding to the Nyquist rate. Hence it is challenging to achieve such stringent requirements of accurate and dynamic delays, especially the signal bandwidth is large [155].

2.5.3 Modulated Wideband Converter

The MWC architecture [65] has been designed in a way resembling both of random demodulator and multicoset sampler - it has multiple branches, and each branch is implemented in a similar manner as the random demodulator to mix with a specific pseudorandom sequence and pass an anti-aliasing lowpass filter [73]. The schematic dia-

Part Number	FPBW (typ) Hz	Resolution bits	Sample Rate (max) Hz	FPBW/Sample Rate	ADC SNR in dBFS (typ) dBFS	Power (typ) W	Price $ US
AD9689-2000	9G	14	2G	4.5	62.7	2.5	$934.86
AD9689-2600	9G	14	2.6G	3.46	61.3	3.1	$952.00
AD9208	9G	14	3G	3	60.2	3.3	$1,190.00
AD9213	6.5G	12	10G	0.65	57.5	5.1	$1,428.00
AD9625-2000	3.2G	12	2G	1.6	59.5	3.48	$624.75
AD9625-2500	3.2G	12	2.5G	1.28	58.3	3.9	$918.74
AD9625-2600	3.2G	12	2.6G	1.23	58.1	4	$1,046.78
AD9697	2G	14	1.3G	1.54	65.6	1.01	$415.65
AD9695-1300	2G	14	1.3G	1.54	65.6	1.6	$760.75
AD9695-625	2G	14	625M	3.2	67.2	980m	$335.75
AD9684	2G	14	500M	4	69.2	2.2	$272.00
AD9690-1000	2G	14	1G	2	67.2	2	$292.19
AD9690-500	2G	14	500M	4	69.2	1.5	$136.00
AD9234-1000	2G	12	1G	2	64.2	3	$389.90
AD9234-500	2G	12	500M	4	65.9	2.15	$238.00
AD9680-1000	2G	14	1G	2	67.2	3.3	$584.38
AD9680-1250	2G	14	1.25G	1.6	63.6	3.7	$692.75
AD9680-500	2G	14	500M	4	69.2	2.2	$272.00
AD9680-820	2G	14	820M	2.44	67.2	2.9	$369.75
AD9691	1.5G	14	1250M	1.2	63.4	3.8	$692.75
AD9694	1.4G	14	500M	2.8	67.1	1.66	$454.75
AD9434-500	1G	12	500M	2	65.9	690m	$124.95
AD9484	1G	8	500M	2	47	670m	$36.00
AD9286	700M	8	500M	1.4	49.3	315m	$36.00

Figure 2.11: Anglog Devices wideband ADC ($\geq$ 500 MHz) product range (information acquired on 16/12/2020).

gram of MWC architecture is illustrated in Fig. 2.12. The design of MWC, although seemingly more complex in hardware terms, avoids some of the practical problems of both random demodulator and multi-coset sampler. It can achieve any fractional compression rate, and removes the need for dynamic and accurate delay distributions required by multicoset samplers. Also, it worth noting that the MWC architecture has filters designed to have anti-aliasing lowpass frequency response to limit signal bandwidth seen by the ADCs, which is only a fraction of the Nyquist rate. Different from the multicoset sampler, the introduction of the pseudorandom mixing and the anti-aliasing filters in MWC effectively alleviates the requirement of a large FPBW of the ADC devices [65].

Denote the multiband time-domain signal as $s(t)$ bandlimited to $\left[-\frac{B}{2}\text{ (Hz)}, \frac{B}{2}\text{ (Hz)}\right]$ where there exists C channels with equal bandwidth of $\frac{B}{C}$ (Hz). The periodic pseudo-random sequences have the chip rate of $1/T_s = B$ (Hz) and period of T_s, formulated as

$$p_p(t) = \sum_{i=0}^{N-1} \beta_{p,i} \text{rect}_{T_s}(t - i \cdot T_s), \quad 0 \leq t < CT_s, \tag{2.66}$$

where the chip waveform is rectangular as proposed in [65]; $p_p(t)$ is the sequence of

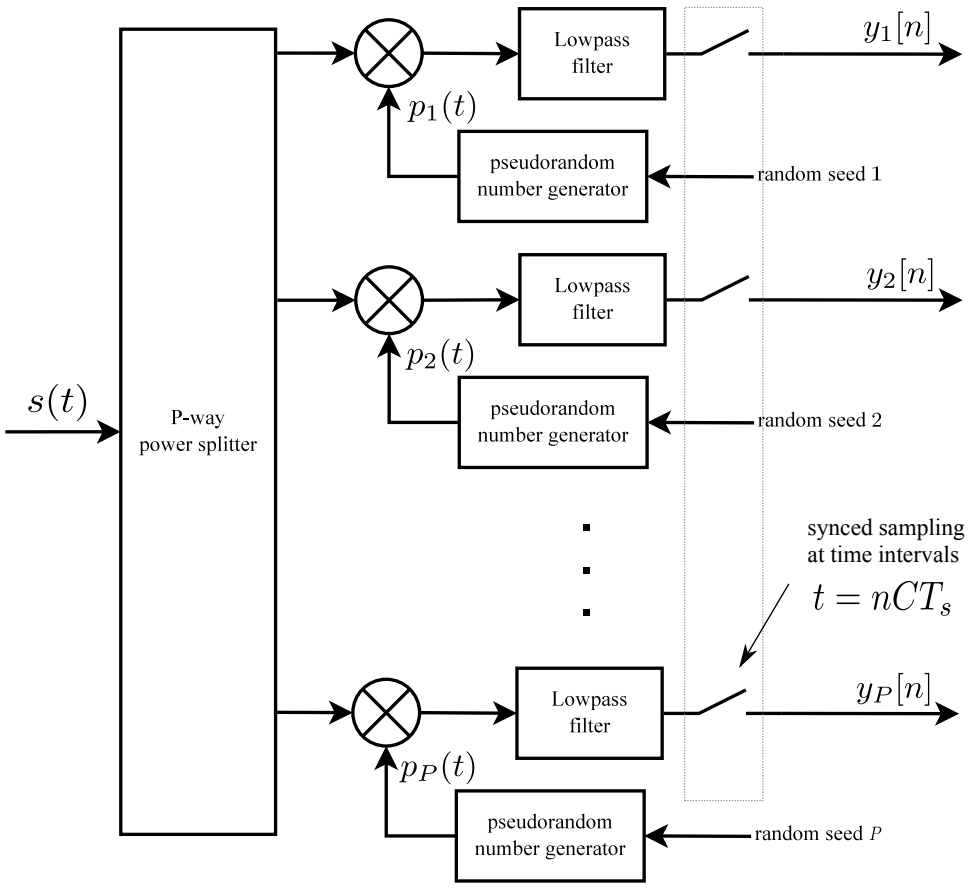


Figure 2.12: MWC architecture.

the pth branch with $\beta_{p,i}$ being the amplitudes of the ith chip in the pth branch. The following low-pass filters are supposed to have identical impulse response with a passband of $\left[0, \frac{B}{C}\ (\text{Hz})\right]$. Then, the ADC array samples the signals of all branches in a synchronized fashion at sub-Nyquist rate of $\frac{B}{C}$ (Hz).

Similar to the multicoset sampler, the intuitions of the signal processing procedures of the MWC can be viewed in the frequency domain. Fig. 2.13 (a) exemplifies the Fourier transform of a $\frac{1}{T_s}$-chip-rate pseudorandom sequence, which is an impulse train with amplitudes being the Fourier series coefficients, determined jointly the by the chip sequence $\beta_{p,i}$'s ($i = 0, 1, \cdots, C-1$) and the chips' waveform. The mixing process can be viewed as the convolution of the spectrum of input signal $s(t)$, shown in Fig. 2.13 (b), with the frequency-domain impulse train in Fig. 2.13 (b), resulting in the frequency

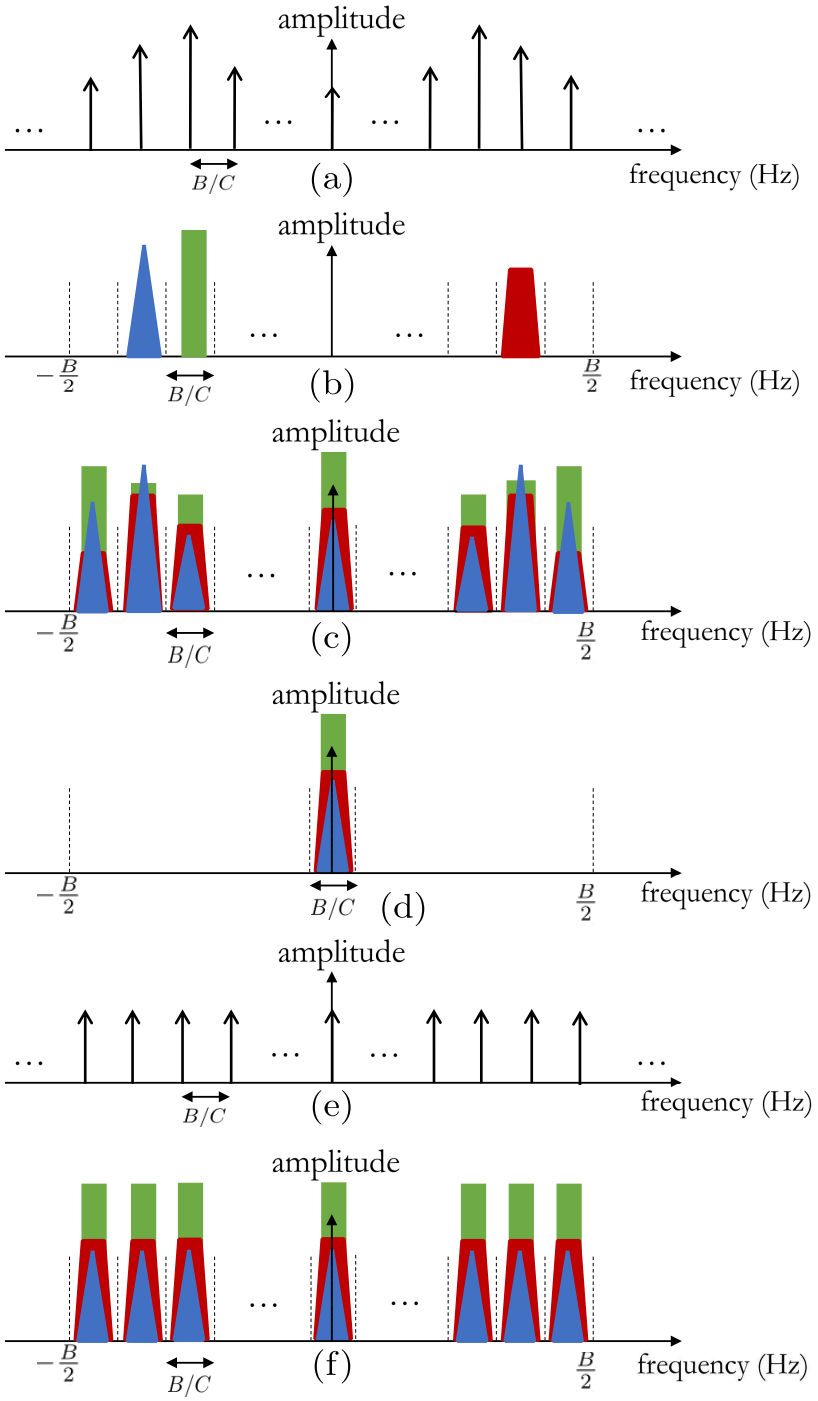


Figure 2.13: Frequency-domain illustration of the sampling process of the MWC. **(a)** The Fourier transform of the pseudorandom sequence $p_p(t)$; **(b)** the spectrum of the input multiband signal $s(t)$ with channel sparsely occupied; **(c)** the Fourier transform of the mixer's output $s(t)p_p(t)$, as the convolution of (a) and (b); **(d)** the Fourier transform of the low-pass filter's output with the passband of $\left[0, \frac{B}{C}\ (\text{Hz})\right]$; **(e)** the Fourier transform of the sampling pulse train; **(f)** the Fourier transform of the discrete-time samples $y_p[n]$, $n \in \mathbb{Z}$, as the convolution of (d) and (e).

representation in Fig. 2.13 (c), effectively creating linear combinations of the spectrum in each of the C channels using the Fourier series coefficients. The following low-pass filter preserves only the linear combination of the spectrum within the frequency range of $\left[0, \frac{B}{C}\ (\text{Hz})\right]$, as shown in Fig. 2.13 (d). Next, the ADC samples this filtered output with the sub-Nyquist rate of $\frac{B}{C}$ (Hz), whose frequency-domain interpretation is the convolution of the signal in Fig. 2.13 (d) with the Fourier transform of the sampling pulse train in Fig. 2.13 (e). The final discrete-time samples $y_p[n]$'s ($n \in \mathbb{Z}$) has the frequency-domain representation shown in Fig. 2.13 (f).

Following the derivations from Mishali et. al. [65], the mathematical model of MWC can be written as

$$\boldsymbol{y}(f) = \boldsymbol{HFDx}(f) = \begin{bmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,C-1} \\ \beta_{2,0} & \beta_{2,1} & \cdots & \beta_{2,C-1} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{P,0} & \beta_{P,1} & \cdots & \beta_{P,C-1} \end{bmatrix} \boldsymbol{FDx}(f) = \boldsymbol{Ax}(f), \tag{2.67}$$

where one can find that pth row of $\boldsymbol{H}$ corresponds to the coefficients of the pseudorandom sequence in the pth branch. The measurement vector

$$\boldsymbol{y}(f) := \left[Y_1\left(e^{j2\pi f\frac{C}{B}}\right), Y_2\left(e^{j2\pi f\frac{C}{B}}\right), \cdots, Y_P\left(e^{j2\pi f\frac{C}{B}}\right)\right]^T$$

consists of p DTFT coefficients of the samples from the pth channel, where $Y_p\left(e^{j2\pi f\frac{C}{B}}\right)$ denotes the DTFT of the discrete samples $y_p[n]$'s ($n \in \mathbb{Z}$). The vector

$$\boldsymbol{x}(f) := \left[X\left(f - \frac{B}{2}\right), X\left(f - \frac{B}{2} + \frac{B}{C}\right), \cdots, X\left(f + \frac{B}{2} - \frac{B}{C}\right)\right]^T$$

contains frequency-shifted versions of the spectrum $X(f)$ of the input signal $s(t)$, representing the spectrum from each of the C channels. $\boldsymbol{F}_{C\times C}$ is a permuted C-order IDFT matrix whose entries are $\{\boldsymbol{F}\}_{p,q} = \exp\left[j2\pi p\left(q - 1 - \frac{C}{2}\right)\right]$. The diagonal matrix has the definition of $\boldsymbol{D} := \text{diag}\left[d_{-\frac{C}{2}}, d_{-\frac{C}{2}+1}, \cdots, d_{\frac{C}{2}}\right]$, and $d_l = \frac{1-\exp\left(-j2\pi\frac{l}{C}\right)}{j2\pi l}$. The sens-

ing matrix $\boldsymbol{A} := \boldsymbol{HFD}$ multiplying the vector $\boldsymbol{x}(f)$ can be understood to describe the frequency-domain convolution process between the input signal $s(t)$ with the pseudorandom sequences $p_p(t)$'s, where specifically each row of $\boldsymbol{A}$ corresponds to the Fourier series coefficients of the pseudorandom sequences, and $\boldsymbol{D}$ accounts for the impact of the pseudorandom sequences' (rectangular) waveform on these pseudorandom sequences' Fourier series coefficients. Again, same as the multicoset sampler, the MWC expression (2.67) forms an IMV model, and can be transformed to an MMV model where the signal to recover is a matrix consisting of $\boldsymbol{x}(f)$'s with finite values of f's as columns. Such a matrix is joint (row) sparse if a large number of channels are vacant, which can be recovered by simultaneous sparse recovery algorithms [45, 152–154].

It is noted that in MWC the matrix $\boldsymbol{H}$ has full pseudorandom entries. By designing the distribution of these random sequences one can achieve desired RIP properties of the sensing matrix, i.e. $\boldsymbol{A} = \boldsymbol{HFD}$, leading to more robustness and compressibility in CS recovery. Gold sequences with binary amplitudes $\beta_{p,i}$ are proposed and the corresponding CS performance is analyzed [156]. Compared with the multicoset sampler, the MWC achieves the linear combination of the spectrum of all C channels by the introduction of the mixing stage in the analog domain before the ADCs, rather than at the sampling stage for the multicoset sampler case. Such mixing stage of the MWC enabling the combination of channels' spectrum makes it possible to limit the signal bandwidth to B/C (Hz) by low-pass filtering at the front end of the ADCs. Thus, it is a major advantage that the MWC structure does not require the ADCs to have the FPBW of B (Hz), but only a fractional of it, i.e. B/C (Hz) [65]. However, the Nyquist chip rate and the synchronization of the pseudorandom sequences, along with the non-ideal low-pass cutoff and inconsistency of the filters are the challenges of its real-world implementations.

2.5.4 Noise in the Signal Acquisition of Compressive Sensing

It is vital to understand the compressive sampler's performance against distortions, which affects the SNR of the acquired signal or, in other words, the dynamic range

of the receiver. Similar to the conventional non-compressive signal acquisition, in the case of compressive sampling, noise appearing in the front end of the analog-digital conversion and the distortions at the conversion process jointly contribute to the noise in the discrete-time samples, i.e. the compressed measurements. Furthermore, the noise in the compressive measurements leads to impairments in the recovered signal. Analytical examinations are necessary to reveal the implications of compressive sampling and CS signal reconstruction on the signal-to-noise performance of the recovered signal. Starting with the following model slightly different from the CS model (2.7),

$$\boldsymbol{y}_{M\times 1} = \boldsymbol{A}_{M\times N}\left(\boldsymbol{x}_{N\times 1} + \boldsymbol{n}_{N\times 1}\right),\ M < N, \tag{2.68}$$

the noise term $\boldsymbol{n}$ represents the noise additive to the Nyquist-rate sparse signal $\boldsymbol{x}$ appearing at the front end of the analog-to-digital conversion, and $\boldsymbol{x}$ is the Nyquist-rate sparse signal to recover. To provide a conventional signal acquisition counterpart to compare with the compressive sampling case, the signal acquired by conventional sampling is thought to be the original Nyquist rate samples with additive noise, i.e. $\boldsymbol{x} + \boldsymbol{n}$. Apart from the additive noise, the quantization at the analog-to-digital conversion further contributes to the error at the measurements. For the additive-noise-free case, the quantized measurements by a n-bit quantization function are expressed by $\mathcal{Q}_n(\boldsymbol{y}) = \mathcal{Q}_n(\boldsymbol{A}\boldsymbol{x})$ and $\mathcal{Q}_n(\boldsymbol{x})$, respectively for compressive sampling and conventional Nyquist-rate sampling. To investigate the impacts of either the additive noise or the quantization noise on the sampling outputs, the SNRs at the output are reviewed separately for each type of error source. Formally, define the input (in-band) SNR (ISNR) as ISNR $:= \frac{N||\boldsymbol{x}||_2^2}{k||\boldsymbol{n}||_2^2}$ and for the compressive sampling case particularly, the recovered SNR (RSNR) as $RSNR := \frac{||\boldsymbol{x}-\boldsymbol{x}^r||_2^2}{||\boldsymbol{x}||_2^2}$, where $k := ||\boldsymbol{x}||_0$ is the sparsity and $\boldsymbol{x}^r$ is the recovered signal from the measurements with the consideration of either additive noise, $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{n})$, or quantization noise, $\mathcal{Q}_n(\boldsymbol{A}\boldsymbol{x})$.

2.5.4.1 Additive Noise

For evaluating the practical performance of radio receivers, the additive noise is commonly measured by the thermal noise floor and the cascaded noise figure of all stages of front-end components. Before discussing the noise performance of the sampling stage, it should be noted firstly that some additional hardware in the compressive sampling architectures may worsen the total front-end noise figure compared to direct Nyquist-rate sampling. For the multicoset sampler and MWC, for instance, a P-way power splitter is necessary which inevitably introduces a noise figure of at least $10\log_{10} P$ dB. Moreover, the noise figure of the mixer output in the random demodulator and MWC may be affected by the noise in the pseudorandom sequences and multiplication imperfections.

Using the RIP of the sensing matrix $\boldsymbol{A}$ with orthogonal rows and with the assumption of using a greedy algorithm to find the correct support of the signal in question $\boldsymbol{x}$, i.e. $\text{supp}(\boldsymbol{s}^r) = \text{supp}(\boldsymbol{s})$, it is shown [83] that the relationship between the ISNR and RSNR is bounded by the following inequalities

$$\frac{M(1-\delta_k)}{N} \leq \frac{\text{RSNR}}{\text{ISNR}} \leq \frac{M(1+\delta_k)}{N}, \tag{2.69}$$

where δ_k is the RIP constant of sparsity k. The RSNR-ISNR gain in (2.69) appears to be the 'best-case' as this gain is expected to be less for the 1-norm minimization sparse recovery, where it is implied that for 1-norm minimization recovery [83]

$$\frac{k(1-\delta_k)}{N(1+\delta_k)} \lesssim \frac{\text{RSNR}}{\text{ISNR}} \lesssim \frac{k(1+\delta_k)}{N(1-\delta_k)}. \tag{2.70}$$

The superiority in terms of RSNR of greedy algorithms over 1-norm minimization can be interpreted by the fact that a greedy algorithm automatically zeros the elements in $\boldsymbol{x}^r$ that are not detected as the support, while for 1-norm minimization noise terms are present in every element of $\boldsymbol{x}^r$. Hence, a factor of $\frac{k}{M}$ of the bounds in (2.69) over (2.70) can be generally expected. Consider a small δ_k, one can conclude a 'best-case'

'rule-of-thumb' gain of SNR due to the additive noise,

$$\left(\frac{\text{RSNR}}{\text{ISNR}}\right)_{\text{dB}} \approx 10\log_{10}\left(\frac{M}{N}\right). \tag{2.71}$$

The conventional Nyquist-rate sampling of the signal $\boldsymbol{x}+\boldsymbol{n}$ preserves the ISNR. Moreover, a commonly-used oversampling technique, consisting super-Nyquist sampling and decimation to recover the resultant Nyquist-rate signal, would lead to an SNR gain of $10\log_{10}(r)$ where r denotes the oversampling rate, i.e. the ratio of the actual sampling rate over the Nyquist rate [157]. The same '3dB SNR gain per octave' rule also applies for undersampling schemes, such as band pass sampling [110], known as 'noise folding' for the interpretation in the frequency domain [84]. This result somewhat coincides with the general rule (2.71), where the oversampling rate can be thought effectively as $M/N < 1$, i.e. the compression ratio. To summarize, compressive sampling implies a '3dB per half of the sampling rate' deterioration of SNR due to additive noise, which poses a significant challenge in the dynamic range of the receiver and hence the implementation of real-world compressive sampling systems.

2.5.4.2 Quantization Noise

In conventional non-compressive signal acquisition, the measurement-to-quantization-noise ratio follows definition and expression of

$$\text{MQNR}_{\text{NYQ}}(\boldsymbol{x}) := \frac{||\mathcal{Q}_n(\boldsymbol{x})-\boldsymbol{x}||_2^2}{||\boldsymbol{x}||_2^2} \geq \left(\frac{2^n}{\gamma(\boldsymbol{x})}\right)^2, \tag{2.72}$$

where n is the number of bits of the midrise uniform quantization function and the signal $\boldsymbol{x}$ is assumed to marginally saturate the quantizer [83]. The term $\gamma(\boldsymbol{x}) := \frac{||\boldsymbol{x}||_\infty \cdot \sqrt{N}}{||\boldsymbol{x}||_2}$ is referred to as the peak-to-average ratio (PAR) of $\boldsymbol{x}$, which measures the ratio between the signal's peak value and power. The MQNR in (2.72) translates to the well-known '6dB per bit' relationship of the quantization SNR, $\left(\text{MQNR}_{\text{NYQ}}(\boldsymbol{x})\right)_{\text{dB}} \gtrsim 6.02n - 20\log\left(\gamma(\boldsymbol{x})\right)$. In the compressive sampling case, the MQNR of the measurements $\mathcal{Q}_n\left(\boldsymbol{y}\right) = \mathcal{Q}_n\left(\boldsymbol{A}\boldsymbol{x}\right)$

can be analyzed by the bound [83]

$$(\mathrm{MQNR}_{\mathrm{CS}}(\boldsymbol{y}))_{\mathrm{dB}} \gtrsim 6.02n + 20\log\left(\frac{\sqrt{(1-\delta_k)\frac{N}{M}}||\boldsymbol{x}||_\infty}{\gamma(\boldsymbol{x})||\boldsymbol{y}||_\infty}\right). \tag{2.73}$$

Furthermore, by introducing a recovery guarantee of the recovery algorithm as $||\boldsymbol{x}^r - \boldsymbol{x}||_2 \leq \kappa||\mathcal{Q}_n(\boldsymbol{y}) - \boldsymbol{y}||_2$, the relationship of the quantization-induced RSNR with the MQNR is found as $\mathrm{RSNR} \geq \frac{\mathrm{MQNR}_{\mathrm{CS}}(\boldsymbol{y})}{(1+\delta_k)\kappa}$, and consequently the quantization-induced RSNR can be expressed in decibels as

$$\mathrm{RSNR} \gtrsim 6.02n + 20\log_{10}\left(\frac{\sqrt{(1-\delta_k)\frac{N}{M}}||\boldsymbol{x}||_\infty}{\gamma(\boldsymbol{x})||\boldsymbol{y}||_\infty}\right) - 20\log_{10}\left((1+\delta_k)\kappa\right), \tag{2.74}$$

where $\kappa > 1$ is a constant in the recovery guarantee [83]. By comparing the RSNR of the compressive sampling (2.74) against the MQNR of conventional Nyquist-rate sampling (2.73), other than the small constant terms δ_k and κ, the main difference lies in the additional coefficient on the PAR-related term, which writes $\frac{\sqrt{\frac{N}{M}}||\boldsymbol{x}||_\infty}{||\boldsymbol{y}||_\infty}$. The coefficient contains both $\boldsymbol{x}$ and $\boldsymbol{y}$, implying a connection to the sensing matrix $\boldsymbol{A}$. By assuming a sub-Gaussian random matrix $\boldsymbol{A}$, such coefficient can be concluded to have a bound with a high probability, i.e. $\frac{\frac{N}{M}||\boldsymbol{x}||_\infty^2}{||\boldsymbol{y}||_\infty^2} \geq \frac{\gamma^2(\boldsymbol{x})}{4\log M}$, further making the bound in (2.74) to finally become

$$\mathrm{RSNR} \gtrsim 6.02n - 20\log_{10}\left(2\sqrt{\frac{\log M}{1-\delta_k}}\right) - 20\log_{10}\left((1+\delta_k)\kappa\right). \tag{2.75}$$

with high probability [83]. Note that the bound in (2.75) is irrelevant of the PAR, indicating that in compressive sampling, the RSNR of the recovered signal can be improved against $\mathrm{MQNR}_{\mathrm{NYQ}}$ for an input signal $\boldsymbol{x}$ with large PAR. By comparing (2.75) and (2.73), the RSNR's gain (in dB) over $\mathrm{MQNR}_{\mathrm{NYQ}}$ is approximated by $\left(\frac{\mathrm{RSNR}}{\mathrm{MQNR}_{\mathrm{NYQ}}}\right)_{\mathrm{dB}} \approx 10\log_{10}\left(\frac{\gamma^2(\boldsymbol{x})}{4\log M}\right)$, which is in the range of

$$10\log_{10}\left(\frac{1}{4\log M}\right) \lesssim \left(\frac{\mathrm{RSNR}}{\mathrm{MQNR}}\right)_{\mathrm{dB}} \lesssim 10\log_{10}\left(\frac{N}{4\log M}\right), \tag{2.76}$$

based on the PAR range of the signal $\boldsymbol{x}$, that is $1 \leq \gamma^2(\boldsymbol{x}) \leq N$ [83]. Looking at such bounds in (2.76), one can expect that the RSNR induced by quantization error to be (1) greatly improved from the MQNR if the input signal has high PAR and (2) deteriorated if the input signal has low PAR. Such potential improvement of the quantization-induced SNR can be explained intuitively by the 'democratic' sampling process in the CS dictated by $\boldsymbol{A}$, where elements in $\boldsymbol{y}$ tend to have 'equal weights' of all the elements in $\boldsymbol{x}$ and thus a high-PAR input $\boldsymbol{x}$ can be transformed into a low-PAR measurement vector $\boldsymbol{y}$.

It should be pointed out that in the signal models used in the remaining of the thesis, quantization distortions are not modeled specifically and the quantization noise can be generally thought of as consisting of part of the additive noise in the measurements, contributing to the SNR of the corresponding CS signal model.

2.6 Summary

This chapter covers the literature review of cognitive radio, dynamic spectrum sharing and conventional and sub-Nyquist wideband spectrum sensing techniques. Background of sampling theories, compressive sensing theories and related techniques for sub-Nyquist wideband spectrum sensing, including sparse recovery algorithms and sub-Nyquist sampling architecture is also presented. As already highlighted in Chapter 1, there are a few major challenges in CSS which lead to the remaining chapters of this thesis.

Chapter 3

Low-Complexity Compressive Spectrum Sensing with Sparsity Estimation

Compressive sensing (CS) techniques have been proposed for wideband spectrum sensing applications to achieve sub-Nyquist-rate sampling. The complexity of the CS recovery algorithm and the detection performance against noise are two of the main challenges of the implementation of compressive spectrum sensing (CSS). Greedy algorithms have been of particular interest in CSS due to their low complexity. A novel spectrum sparsity estimation scheme directly from sub-Nyquist measurements is firstly proposed, with which the computational effort of greedy pursuit algorithms can be saved and recovery performance improved. Besides, the spectrum sparsity estimates also enable hard detection of channel occupancy where threshold adaption for energy detection is avoided. Moreover, with the detected dimension of signal subspace, it is proposed to implement a joint-block-sparse multiple-measurement-vector (MMV) model of CSS whose dimension can be reduced to the minimum, and meanwhile, a large portion of the noise is removed. The proposed MMV model with noise and dimension reduction further improves the

detection performance and also keeps the complexity low. Finally, the hard thresholding pursuit algorithm is generalized to recover joint-block-sparse signals. In simulations, the detection performance and complexity of the proposed CSS scheme show evident superiority against multiple benchmarking schemes. The rest of this chapter is structured as follows. Section 3.1 gives the introduction of the related work and the outlined contribution of the work in this chapter. In Section 3.2, system and signal models of CSS are presented. In Section 3.3, the proposed spectrum sparsity and active channel number estimation scheme are detailed. In Section 3.4, the dimension reduction of the MMV model and noise removal is introduced. Then, performances of various commonly-used greedy algorithms are discussed and the HTP algorithm is generalized for joint-block-sparse signals. Monte-Carlo simulation results and further comments are given in Section 3.5.

3.1 Introduction

3.1.1 Related Work

Considering the demand for DSA applications, rapid spectrum reconstruction and robust spectrum availability detection are desired in CSS. Optimization methods in CSS have been discussed in literature, including convex l_1 optimization [14, 15], non-convex l_v $(0 < v < 1)$ optimization [53], and sparse Bayesian learning [94, 158]. However, the complexity of CS recovery algorithms, especially these optimization methods, has been one of the major bottlenecks of the implementation of CSS applications. Among these algorithms, greedy algorithms, especially the Hard Thresholding Pursuit (HTP) [149] and its derivations, have drawn great attention due to their lower computational cost compared to these optimization-based sparse recovery schemes [55]. Most greedy algorithms for sparse recovery require the sparsity order of the signal to be recovered as a priori input information. Inputting an excessively large sparsity tends to lead to unnecessary free dimensions of the signal space, which does not help detect the support of the desired signal and makes the recovery fidelity worse. However, in practice, valid estimates of

sparsity in a dynamic scenario are difficult to obtain, hence the *a priori* known upper bound of the sparsity based on long-term observations of the spectrum of interest or the maximum allowed value of sparsity is often input to these algorithms to ensure reliable spectrum reconstruction. Another simple solution is to directly use the channel sparsity information from the geo-location database. However, such information could be inaccurate as geo-location database relies on a propagation model which is designed to give conservative channel availability to protect PUs [6, 45]. Some spectrum sparsity estimation solutions have also been seen in the literature [58, 59]. In these works, prior to the actual spectrum sensing step, the spectrum sparsity is estimated offline using the recovered signals from multiple trials of recovery implementations, which makes the actual spectrum sensing inefficient. In [90], a novel metric for sparsity is proposed for noisy signals, and estimation is performed directly from compressive measurements. However, the measurements are required to be sampled by various measurement matrices of a special formation. Moreover, the sparsity information in [58, 59, 90] is proposed to be applied to convex-optimization and adaptively optimize the compression ratio, which is fundamentally different from how the sparsity estimation is proposed to be used in this chapter.

Robust detection of spectrum availability is also a crucial procedure in CSS. The simple yet commonly adopted method is Neyman-Pearson (NP) energy detection (ED) on each channel's spectrum energy, which determines the channel occupancy in a soft decision manner by setting a proper threshold [103]. Some work in CSS proposes cyclostationary feature extraction on the recovered spectrum and then performs ED [158, 159]. Some prior information on the noise statistics is essential to optimal threshold setting. Practical noise estimation and threshold setting schemes for conventional non-compressive spectrum sensing have been proposed in the literature [97, 160]. However, it is discovered in [161] that the statistical model of the channel energy based on the recovered spectrum differs from that of the conventional model, and the parameters of these models have to be learned from multiple recovered spectrum samples in order to set the proper

threshold. The proposed parameter learning scheme also has the realistic shortcoming that the estimation requires time-consuming recovery operations.

3.1.2 Contributions

To address the sparsity estimation issue, a novel sparsity estimation directly from time-domain sub-Nyquist samples is presented. The proposed sparsity estimation is achieved by the subspace analysis of the auto-correlation matrix of sub-Nyquist samples. In the area of array processing, a similar problem is source enumeration. The full-row-rank 'sensing matrix' in the CSS context is an analogy of the full-column-rank 'array steering matrix', and the spectrum sparsity is actually equivalent to the number of sources in array processing. Source numeration problem has been being studied for decades, with some information theoretic criteria (ITC) proposed [162, 163] to select the optimal number of principle and noise subspace dimensions in the sense of information distance through eigendecomposition, including Akaike's information criterion [164] and minimum description length (MDL) [165] (or equivalently Bayesian Information Criterion (BIC) in Bayesian formulation [166]). BIC has drawn particular interests due to its better asymptotic performances [167]. It is chosen to use an enhanced BIC strategy to achieve the spectrum sparsity estimation. A pioneering subspace-aided sparsity estimation strategy has been recently outlined in [16]. However, the specific estimation methods applied, estimation performance, or the advantages of spectrum sparsity estimates have not been discussed.

This chapter considers a general multi-band signal model where the baseband spectrum consists of multiple channels of equal bandwidth and only sporadic channels have active signals present. In the CS model, such signal's spectrum to be reconstructed has a joint-sparse structure. The proposed spectrum sparsity estimation is based on multiple samples of the sub-Nyquist measurement vectors and the sparsity estimate can be translated to the estimate of active channel numbers. Note that the proposed spectrum sparsity estimation scheme does not assume hence is not limited to a joint-sparse spectrum

model, and therefore, for some frequency-sparse signals that do not have sparse channel occupancy, the proposed spectrum sparsity estimation still applies to obtain a sparsity input for greedy algorithms for general sparse signals. Next, it is further proposed to use the multiple-measurement-vector (MMV) model for joint-block-sparse signals for CS directly instead of the single-measurement-vector (SMV) model. Based on the subspace decomposition of the sparsity estimation scheme, it is natural to reduce the dimensionality of the MMV model to its minimum by removing the noise subspace, which effectively achieves denoising from the measurements. Finally, the HTP algorithm is extended for recovering joint-block-sparse signals, which empirically gives better detection performance than l_1 optimization and orthogonal matching pursuit (OMP) counterparts while keeping the complexity low.

The merits of the proposed spectrum sparsity estimation scheme and the implementation of dimension-reduced MMV model are summarized as follows:

- The estimation of spectrum sparsity is directly from compressed measurements, which does not require recovery operations;
- A valid estimate of spectrum sparsity aids the greedy pursuit algorithm to reduce the complexity and enhance the recovery performance;
- Dimension reduction of the MMV model further improves the recovery performance and reduces recovery complexity.
- With a valid estimate of spectrum sparsity, hard detection (HD) of channel occupancy with the detected number of occupied channels can be achieved without the need of dictating or adapting the threshold.

3.2 System and Signal Models

3.2.1 System Model

Consider a cognitive radio system with primary users (PUs) being 'virtual' transceiver stations that transmit orthogonal frequency division multiplexing (OFDM) signals within C frequency channels within a band whose bandwidth is B (Hz). Each channel has an equal bandwidth of B/C (Hz). The OFDM signal model is straightforward to be used to formulate a multiband signal with arbitrary frequency supports defined by the OFDM symbols carried by a series of monotone sub-carriers. The OFDM signal model is therefore adopted purely to conveniently describe a general multiband signal with sparse channel occupancy in the frequency domain, for example as illustrated in Fig. 3.2, where F_s (Hz) is the frequency spacing of the adjacent sub-carriers, in other words, the symbol rate of the imaginary OFDM transmitters. In the upper half in Fig. 3.1, the model diagram of a virtual OFDM-based radio transmitter is illustrated. The baseband signal from the cth transmitter is to be upconverted by $F_0+(c-1)B/C$ (Hz) to the cth channel of the interested wideband spectrum, where the frequency range is $[F_0$ (Hz), F_0+B (Hz)]. Then, the number of sub-carriers for each PU transmitter is determined by B/F_s (Hz), which is assumed to be an integer.

3.2.2 Signal Model

Here, the model of the signal to be sensed is constructed. Firstly, with accordance to the PU transmitter diagram in Fig. 3.1, one can write the continuous time-domain signal of OFDM transmitters within one symbol duration as

$$s_o(t) = \sum_{c=1}^{C} \sum_{d=1}^{\frac{B}{CF_s}} (a_{c,d} + jb_{c,d}) e^{j2\pi[(d-1)F_s+(c-1)B/C]t}, \tag{3.1}$$

where $0 \leq t < (1/F_s)$. $a_{c,d}$ and $b_{c,d}$ represents the dth data symbol from the cth PU, of in-phase and quadrature component respectively. Practical implementations of OFDM are digital via the inverse fast Fourier transform (IFFT) algorithm of B/F_s points, and

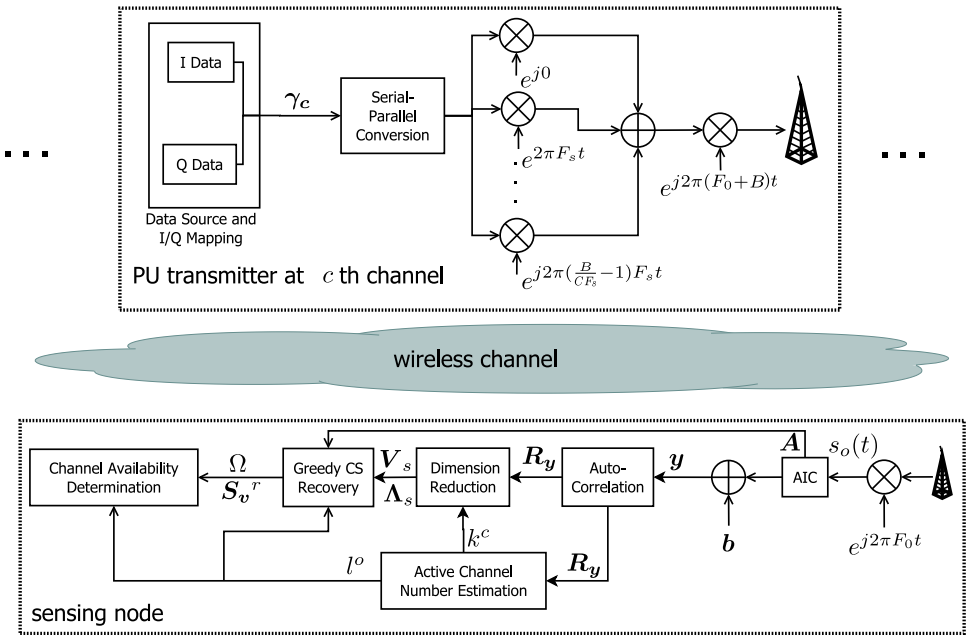


Figure 3.1: The proposed system architecture: PU transmitters and a sensing node. Specific definitions of the signal notations are throughout Chapter 3.

for a wideband system compromising C channel the Nyquist frequency is $f_{\text{NYQ}} = B$ (Hz). The discrete time-domain signal $(\boldsymbol{s_o})_{\frac{B}{F_s}\times 1}$ is equivalent to the continuous version (3.1) sampled at Nyquist frequency f_{NYQ}, that is

$$\boldsymbol{s_o} = \left[s_o(0)\ s_o\left(\frac{1}{B}\right)\ \cdots\ s_o\left(\frac{\frac{B}{F_s}-1}{B}\right)\right]^T = \boldsymbol{G\gamma}. \tag{3.2}$$

where $\boldsymbol{G}_{\frac{B}{F_s}\times\frac{B}{F_s}}$ is effectively a $\frac{B}{F_s}$-point IDFT matrix. The vector $\boldsymbol{\gamma}$ is defined by $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}_1^T\ \boldsymbol{\gamma}_2^T\ \cdots\ \boldsymbol{\gamma}_C^T\right]^T$, where data symbol blocks $(\boldsymbol{\gamma}_c)_{\frac{B}{CF_s}\times 1}$ $(c = 1, 2, \cdots, C)$ are presented by

$$\boldsymbol{\gamma}_c = \left[a_{c,1} + jb_{c,1}\ a_{c,2} + jb_{c,2}\ \cdots\ a_{c,B/CF_s} + jb_{c,\frac{B}{CF_s}}\right]^T. \tag{3.3}$$

In order to express the sparse channel occupation, for some $1 \le c \le C$, each $\boldsymbol{\gamma}_c$ is a vector with non-zero entries, which represent the frequency coefficients of the occupied channels, and for other $\boldsymbol{\gamma}_c$'s, $\boldsymbol{\gamma}_c = \boldsymbol{0}$ which models the vacant channels. The frequency representation (i.e. the DFT) of $\boldsymbol{s}_o$ is effectively $\boldsymbol{\gamma} = \boldsymbol{G}^{-1}\boldsymbol{s}_o = \boldsymbol{G}^{-1}\boldsymbol{G\gamma}$. At the receiver's

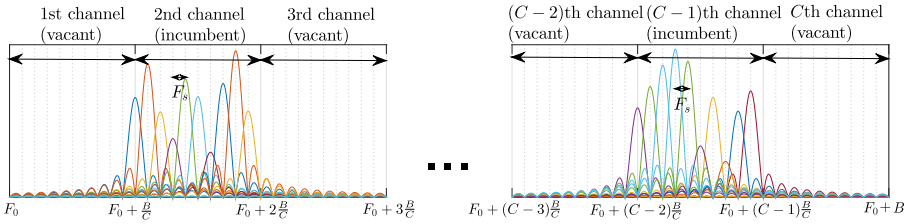


Figure 3.2: A conceptual illustration of the interested TVWS wideband spectrum.

end, the received frequency-domain signal through a generic multipath channel is $\hat{\boldsymbol{\gamma}} = [\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \cdots, \hat{\boldsymbol{\gamma}}_C]$, where

$$\hat{\boldsymbol{\gamma}}_c = L_c \left[\frac{1}{\sqrt{K_c+1}} + \frac{\sqrt{K_c}}{\sqrt{K_c+1}} \cdot \boldsymbol{h}_c \right] \circ \boldsymbol{\gamma}_c, \;\; 1 \leq c \leq C. \tag{3.4}$$

$\circ$ denotes element-wise product. L_c's are the path loss from the transmitter occupying the cth channel and the receiver. $\boldsymbol{h}_c$'s are multipath channel gain where $(\boldsymbol{h}_c)_{\frac{B}{CF_s}\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. K_c is the power ratio between multipath and line-of-sight components. Denote the time-domain representation of the received signal by $\hat{\boldsymbol{s}_o} := \boldsymbol{G}^{-1}\hat{\boldsymbol{\gamma}}$.

Consider a simple single-node non-cooperative CSS scenario. Denote the Nyquist-rate time-domain signal as $\boldsymbol{s}_{tN\times 1} = [s_t[0]\ s_t[1] \ldots s_t[N-1]]^T$ and its frequency-domain representation $\boldsymbol{s}_f = \boldsymbol{\mathcal{F}}\boldsymbol{s}_t = [s_f[0]\ s_f[1] \ldots s_f[N-1]]^T$ where $\boldsymbol{\mathcal{F}}_{N\times N}$ stands for the N-point discrete Fourier transform (DFT) matrix. The sparse signal $\boldsymbol{s}_f$ with the sparsity denoted by $\delta(\boldsymbol{s}_f) := ||\boldsymbol{s}_f||_0$, can be recovered from a M-element sub-Nyquist-rate measurement vector $\boldsymbol{y}_{M\times 1}$ where $M < N$, with the product (i.e. sensing matrix) $\boldsymbol{A} := \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}$ meeting the restricted isometry property [55]. The sub-Nyquist-rate sampling can be characterized by a linear system

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{s}_t + \boldsymbol{b} = \boldsymbol{A}\boldsymbol{s}_f + \boldsymbol{b}, \tag{3.5}$$

where $\boldsymbol{\Phi}_{M\times N}$ is the sampling matrix with structured random entries corresponding to analog-to-information converter (AIC) settings [71], and $\boldsymbol{b}_{M\times 1} \sim \mathcal{CN}\left(\mathbf{0}_{M\times 1}, \sigma^2\mathbf{I}_M\right)$ is a vector of independent and identically distributed (i.i.d.) complex additive white Gaussian

noise present in the measurements. Recovering a sparse $\boldsymbol{s}_f$ from the SMV model with additive noise (3.5) is a typical problem in the CS area. Denote the recovered signal as $\boldsymbol{s}_f^r$. SNR is formally defined by

$$SNR := \frac{\mathrm{E}\left[||\boldsymbol{\Phi}\boldsymbol{s}_t||_2^2\right]}{(M\sigma^2)}. \tag{3.6}$$

Consider that the practically large number of points $\frac{B}{F_s}$ of the spectrum at receiver $\hat{\boldsymbol{\gamma}}$ may provide unnecessarily high resolution and adds to the complexity for CS. Here, the decimated version of the spectrum is considered as the sparse spectrum to be recovered $\boldsymbol{s}_f$ in (3.5), writing

$$s_f[n] := \hat{\gamma}\left[n \cdot \frac{B}{F_s N}\right], \quad n = 0, 1, \cdots, N-1, \tag{3.7}$$

where it is assumed N is chosen to make $\frac{B}{CF_sN}$ an integer. It is straightforward from (3.7) that $\delta(\boldsymbol{s}_f)$ non-zeros entries in $\boldsymbol{s}_f$ are independent zero-mean complex random variables. The Nyquist-rate time-domain samples $\boldsymbol{s}_t$ in the CS model (3.5) is then related to $\hat{\boldsymbol{s}_o}$ by

$$\begin{aligned}
s_t[n] &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} s_f[i] \cdot e^{j2\pi \frac{in}{N}} \\
&= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} \left[\sum_{m=0}^{\frac{B}{F_s}-1} \hat{s_o}[m] \cdot e^{-j2\pi \frac{m\left(i \cdot \frac{B}{F_s N}\right)F_s}{B}} \right] e^{j2\pi \frac{in}{N}} \\
&= \frac{1}{\sqrt{N}} \sum_{m=0}^{\frac{B}{F_s}-1} \hat{s_o}[m] \left[\sum_{i=0}^{N-1} e^{-j2\pi \frac{i(m-n)}{N}} \right] \\
&= \sqrt{N} \sum_{k=0}^{\frac{B}{NF_s}-1} \hat{s_o}[kN+n], \quad n = 0, 1, \cdots, N-1,
\end{aligned} \tag{3.8}$$

where the last equation is based on the following relationship

$$\sum_{i=0}^{N-1} e^{-j2\pi \frac{i(m-n)}{N}} = \begin{cases} N, & m-n = kN \text{ and } k \in \mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases}$$

The expression (3.8) indicates that by considering the decimated version of spectrum, $s_f[n]$'s, in the CS system (3.5), the time domain samples $s[t]$'s in (3.5) should be obtained by summing every N-sample frames in the raw time domain samples $\hat{s_o}[m]$'s, resulting in a N-sample vector $\boldsymbol{s}_t = [s_t[0], s_t[1], \cdots, s_t[N-1]]^T$.

3.3 Spectrum Sparsity Estimation based on Subspace Decomposition

Stack T samples of the measurement vectors obtained from sub-Nyquist sampling as columns of a matrix, $\boldsymbol{Y}$, expressed by

$$\begin{aligned}\boldsymbol{Y} &:= \left[\boldsymbol{y}^{(1)}\, \boldsymbol{y}^{(2)}\, \cdots\, \boldsymbol{y}^{(T)}\right] \\ &= \boldsymbol{A}\left[\boldsymbol{s}_f^{(1)}\, \boldsymbol{s}_f^{(2)}\, \cdots\, \boldsymbol{s}_f^{(T)}\right] + \left[\boldsymbol{b}^{(1)}\, \boldsymbol{b}^{(2)}\, \cdots\, \boldsymbol{b}^{(T)}\right] \\ &= \boldsymbol{A}\boldsymbol{S}_f + \boldsymbol{B} = \boldsymbol{Y}_s + \boldsymbol{B},\end{aligned} \tag{3.9}$$

where $\boldsymbol{S}_f := \left[\boldsymbol{s}_f^{(1)}\, \boldsymbol{s}_f^{(2)}\, \cdots\, \boldsymbol{s}_f^{(T)}\right]$ and $\boldsymbol{B} := \left[\boldsymbol{b}^{(1)}\, \boldsymbol{b}^{(2)}\, \cdots\, \boldsymbol{b}^{(T)}\right]$. The underdetermined system (3.9) consists of a MMV model in the CS literature, where the spectrum support can be determined after recovering the joint row-sparse matrix $(\boldsymbol{S}_f)_{N\times T}$. It is hereby assumed that the channel occupancy remains the same in the period of sample acquisition T/F_s, hence $\boldsymbol{S}_f$ is joint row-sparse. Apart from the joint row-sparse property of $\boldsymbol{S}_f$, it also features a block-sparse property due to the channel structure, where every single $(\frac{N}{C} \times T)$-element block corresponding to the same channel contains non-zero elements or all zero elements for an occupied or vacant channel respectively. The sparsity pattern of $\boldsymbol{S}_f$ is hereby termed by 'joint-block-sparse' as exemplified in Fig. 3.3 and the term 'channel' and 'block' are used interchangeably in the remainder of this chapter.

Examine the auto-correlation matrix of sub-Nyquist rate measurement vectors,

$$\boldsymbol{R}_y = \boldsymbol{A}\boldsymbol{R}_{s_f}(\boldsymbol{\mathcal{F}}^{-1})^H\boldsymbol{\Phi}^H + \sigma^2\mathbf{I}_M, \tag{3.10}$$

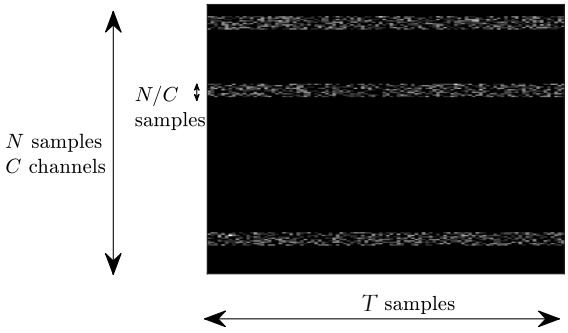


Figure 3.3: A visualization of the joint-block-sparse matrix $\boldsymbol{S}_f$.

where

$$\boldsymbol{R}_y = \mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^H\right] = \lim_{T\to\infty}\frac{1}{T}\sum_{p=1}^{T}\boldsymbol{y}^{(p)}\left(\boldsymbol{y}^{(p)}\right)^H$$

and

$$\boldsymbol{R}_{s_f} = \mathrm{E}\left[\boldsymbol{s}_f\boldsymbol{s}_f^H\right] = \lim_{T\to\infty}\frac{1}{T}\sum_{p=1}^{T}\boldsymbol{s}_f^{(p)}\left(\boldsymbol{s}_f^{(p)}\right)^H.$$

With finite sample size T, one can only obtain estimates of auto-correlation matrices $\boldsymbol{R}_y$ and $\boldsymbol{R}_{s_f}$ denoted by $\hat{\boldsymbol{R}}_y$ and $\hat{\boldsymbol{R}_{s_f}}$, where it holds

$$\begin{aligned}\hat{\boldsymbol{R}}_y =& \boldsymbol{A}\hat{\boldsymbol{R}_{s_f}}(\boldsymbol{\mathcal{F}}^{-1})^H\boldsymbol{\Phi}^H + \frac{1}{T}\boldsymbol{B}\boldsymbol{B}^H \\ =& \frac{1}{T}\boldsymbol{Y}_s\boldsymbol{Y}_s^H + \frac{1}{T}\boldsymbol{B}\boldsymbol{B}^H,\end{aligned} \tag{3.11}$$

where $\hat{\boldsymbol{R}}_y = \frac{1}{T}\sum_{p=1}^{T}\boldsymbol{y}^{(p)}\left(\boldsymbol{y}^{(p)}\right)^H = \frac{1}{T}\boldsymbol{Y}\boldsymbol{Y}^H$, and $\hat{\boldsymbol{R}}_{s_f} = \frac{1}{T}\sum_{p=1}^{T}\boldsymbol{s}_f^{(p)}\left(\boldsymbol{s}_f^{(p)}\right)^H = \frac{1}{T}\boldsymbol{S_f}\boldsymbol{S_f}^H$. In Proposition 1 that follows, it is presented the property on the rank of $\hat{\boldsymbol{R}}_{s_f}$ and $\hat{\boldsymbol{R}}_y$ under noise-free conditions, based on which the sparsity of spectrum is proposed to be estimated.

Lemma 1. *Given $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)} \in \mathbb{C}^r$ is a series of arbitrary vectors and $s < r$, a random vector $\boldsymbol{\alpha}^{(s+1)} \in \mathbb{C}^r$ whose entries are independent and follow an absolutely continuous distribution makes $dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)}, \boldsymbol{\alpha}^{(s+1)}\} = dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)}\} + 1$ with the probability of 1.*

Proposition 1. *Given $\sigma^2 = 0$, and that random vector $\boldsymbol{s}_f$ has the sparsity of $\delta(\boldsymbol{s}_f)$,*

if $T \geq \delta(\boldsymbol{s}_f)$*, it holds with probability of 1 that* $rank(\hat{\boldsymbol{R}_{s_f}}) = \delta(\boldsymbol{s}_f)$*. Furthermore, if the sensing matrix* $\boldsymbol{A}$ *has full rank and* $M \geq \delta(\boldsymbol{s}_f)$*, it holds with probability of 1 that* $rank(\hat{\boldsymbol{R}}_y) = rank(\hat{\boldsymbol{R}_{s_f}}) = \delta(\boldsymbol{s}_f)$*.*

Proofs for Lemma 1 and Proposition 1 are given in the Appendix. Proposition 1 indicates that the sparsity $\delta(\boldsymbol{s}_f)$ can be revealed by the rank of the estimated autocorrelation of the measurements $\boldsymbol{R_y}$ under the conditions that 1) $T \geq \delta(\boldsymbol{s}_f)$, 2) $M \geq \delta(\boldsymbol{s}_f)$ and that 3) the sensing matrix $\boldsymbol{A}$ has full rank. Since CS theory requires the product of the sampling matrix $\boldsymbol{\Phi}$ and the sparsifying matrix $\boldsymbol{\mathcal{F}}^{-1}$ to be nearly orthogonal [55], and the number of measurements to meet $M > \mathcal{O}(\delta(\boldsymbol{s}_f) \log N)$ [168], condition 2) and 3) can be naturally met in most practices of CS. Therefore, one need guarantee that the number of measurement vector samples is larger than the maximum sparsity of the spectrum in order to employ the proposed sparsity estimation scheme.

Performing the eigendecomposition of $\boldsymbol{A}\hat{\boldsymbol{R}_{s_f}}\boldsymbol{A}^H$ which has the rank of $\delta(\boldsymbol{s}_f)$ from Proposition 1 will yield $\delta(\boldsymbol{s}_f)$ non-zero eigenvalues. With the presence of noise, $\hat{\boldsymbol{R}}_y$ is presented by

$$\hat{\boldsymbol{R}}_y = \sum_{i=1}^{\delta(\boldsymbol{s}_f)} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^H + \frac{1}{T}\boldsymbol{B}\boldsymbol{B}^H, \tag{3.12}$$

where $\lambda_1, \cdots, \lambda_{\delta(\boldsymbol{s}_f)}$ denote the (non-zero) eigenvalues in the signal's subspace spanned by the corresponding eigenvectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{\delta(\boldsymbol{s}_f)}$. In practice, the eigendecomposition of estimated auto-correlation matrix $\hat{\boldsymbol{R}}_y$ with the consideration of noise is applied as such

$$\hat{\boldsymbol{R}}_y = \sum_{i=1}^{M} \hat{\lambda}_i \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i^H, \tag{3.13}$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_M$ and $\hat{\boldsymbol{v}}_1, \cdots, \hat{\boldsymbol{v}}_M$ denote the descending-ordered eigenvalues and corresponding eigenvectors of the estimated auto-correlation matrix. The eigenvalue profile based on the autocorrelation estimates $\hat{\boldsymbol{R}}_y$ asymptotically converges to, with a increasing number of observations - T, to the true eigenvalues of $\boldsymbol{R}_y$, which consists of $\delta(\boldsymbol{s}_f)$ dominant eigenvalues corresponding to the signal power components perturbed by

noise and $(M-\delta(\boldsymbol{s}_f))$ small and equal eigenvalues equal to σ^2 only representing the noise power [169]. This implies that the sparsity estimation may be obtained by differentiating two groups of eigenvalues by their amplitudes.

3.3.1 A Bayesian-Information-Criteria-Based Estimator

In practical estimation-based eigendecomposition (3.13), only limited numbers of samples can be used for estimation and the SNR can be low, hence, the noise subspace eigenvalues are not identical and can be difficult to be discriminated from signal subspace eigenvalues. The above-described sparsity estimation problem is analogous to the source enumeration which is a fundamental problem in array processing. A recently-proposed algorithm based on information theoretic criteria (ITC) for estimating the dimensionality of signal subspace is the enhanced Bayesian Information Criterion (BICe) [170] which defines the ITC metric as

$$\begin{aligned}\mathrm{BICe}(k) = &-2\log f(\boldsymbol{Y}|\hat{\lambda}_1,\cdots,\hat{\lambda}_M;k)\\ &-2\log f(\hat{\lambda}_1,\cdots,\hat{\lambda}_M;k)+C_k\log T,\end{aligned} \tag{3.14}$$

$k=1,\cdots,M$,where $f(\cdot)$ denotes probability density function and $C_k = k(2M-k)$ [162] is the number of free parameters related to k. Then the estimation k^o of $\delta(\boldsymbol{s}_f)$ is determined by

$$k^o = \arg\min_k \mathrm{BICe}(k). \tag{3.15}$$

Additionally, note $\boldsymbol{s}_f$ and $\boldsymbol{b}$ are independent and both zero-mean random vectors, one can approximately regard $\boldsymbol{y}$ as a zero-mean Gaussian random vector as a result of central limit theorem. Hence, the posterior probability of T independent observations $\boldsymbol{Y}$, given the estimated eigenvalues parameterized by k, is expressed by the multivariate Gaussian

model [171]

$$\begin{aligned} f(\boldsymbol{Y}|\hat{\lambda}_1,\cdots,\hat{\lambda}_M;k) =& \frac{\prod_{p=1}^{T}\exp\mathrm{tr}\left(\left(\boldsymbol{y}^{(p)}\right)^H\cdot\hat{\boldsymbol{R}}_y\cdot\boldsymbol{y}^{(p)}\right)}{\pi^M\cdot\prod_{i=1}^{k}\hat{\lambda}_i\cdot(\hat{\sigma}^2)^{M-k}} \\ =& \frac{1}{\pi^M\cdot\prod_{i=1}^{k}\hat{\lambda}_i\cdot(\hat{\sigma}^2)^{M-k}}, \end{aligned} \tag{3.16}$$

where $\hat{\lambda}_1,\cdots,\hat{\lambda}_k$ and $\hat{\sigma^2} := \sum_{i=k+1}^{M}\hat{\lambda}_i/(M-k)$ are maximum-likelihood (ML) estimates for $\lambda_1,\cdots,\lambda_k$ and σ^2, respectively.

The joint probability density $f(\hat{\lambda}_1,\cdots,\hat{\lambda}_M;k)$ is proposed [170] to be approximated by the product of the probability density of signal subspace eigenvalues and the probability density of noise subspace eigenvalues, using the ML estimates to approximate the true values, that is

$$\begin{aligned} f(\hat{\lambda}_1,\cdots,\hat{\lambda}_M;k) =& f(\hat{\lambda}_1,\cdots,\hat{\lambda}_k|\lambda_1=\hat{\lambda}_1,\cdots,\lambda_k=\hat{\lambda}_k) \\ & \cdot f(\hat{\lambda}_{k+1},\cdots,\hat{\lambda}_M|\sigma^2=\hat{\sigma}^2). \end{aligned} \tag{3.17}$$

The probability densities $f(\hat{\lambda}_1,\cdots,\hat{\lambda}_k|\lambda_1,\cdots,\lambda_k)$ and $f(\hat{\lambda}_{k+1},\cdots,\hat{\lambda}_M|\sigma^2)$ are derived in [171] and [167] respectively. By relating (3.16) and (3.17) into (3.14), after terms irrelevant to k being removed, the optimization of BICe can be rearranged as the following

$$\begin{aligned} k^o &= \arg\min_k \mathrm{BICe}(k) \\ &= \arg\min_k \Bigg\{ 2(M-k)(T+M-k)\log\hat{\sigma^2} \\ &- 2T\sum_{i=k+1}^{M}\log\hat{\lambda}_i + 2\sum_{i=1}^{k}\log\hat{\lambda}_i + 2\sum_{i=1}^{M-k}\log\Gamma(i) \\ &+ T\sum_{i=k+1}^{M}\left(\frac{\hat{\lambda}_i-\hat{\sigma^2}}{\hat{\sigma^2}}\right)^2 - 4\sum_{i=k+1}^{M}\sum_{j=i+1}^{M}\log|\hat{\lambda}_i-\hat{\lambda}_j| \\ &+ (4Mk-2k^2-k)\log T\Bigg\}, \end{aligned} \tag{3.18}$$

where $\Gamma(\cdot)$ denotes the real Gamma function.

3.3.2 Initial Performance Analysis

As a generalization of the original BIC detector, the BICe-based detector is reported to inherit the *asymptotic consistency property* [170] that both the overestimation and underestimation probability converge to zero with T increasing to infinity, regardless of the SNR. Furthermore, overestimation probability converges much more rapidly, and is reportedly small enough empirically and analytically [172] to be neglected.

The eigenvalue profiles of varying SNR and T are firstly shown in Fig. 3.4 (a) and (b) respectively. With decreasing SNR, it is a natural result that the increasing eigenvalues from noise subspace lead to more difficulty in discriminating the signal subspace eigenvalues from the eigenvalue profile. With decreasing T, the number of noise subspace eigenvalues is also harder to be determined as the noise subspace eigenvalues tend to be more inconsistent in the eigenvalue profile. Moreover, the BICe metrics against $1 \leq k \leq M$ are illustrated with changing SNR and T in Fig. 3.5 (a) and (b). With SNR and T above certain level, the proposed BICe-based detection method can accurately detect the sparsity of spectrum $k^o = \text{rank}(\boldsymbol{R}_{s_f}) = \delta(\boldsymbol{s}_f)$. However, with either further decreasing SNR or T, the k values corresponding to minimum BICe metrics tend to be underestimated, that is $k^o < \delta(\boldsymbol{s}_f)$. Additionally, in Fig. 3.5, overestimation is never spotted with finite times of simulations and the smallest eligible $T = M = 512$. In this sense, one can reasonably match $T = 512$ to the qualitative description 'large sample size' in terms of negligible overestimation probability.

3.3.3 Estimation of Active Channel Number

As the sparse presentation $\boldsymbol{S}_f$ is a joint-block-sparse matrix, having estimated the sparsity of the spectrum, one finally needs the estimated active channel number as the number of blocks to be detected in the recovery algorithm. Hereby, the estimation of active channel number is proposed based on the constraint that the dimension of signal subspace only can be multiples of the number of spectrum bins of each channel, which is $\delta(\boldsymbol{s}_f) = l \cdot N/C$ with the actual number of occupied channels $0 \leq l \leq C$ being some

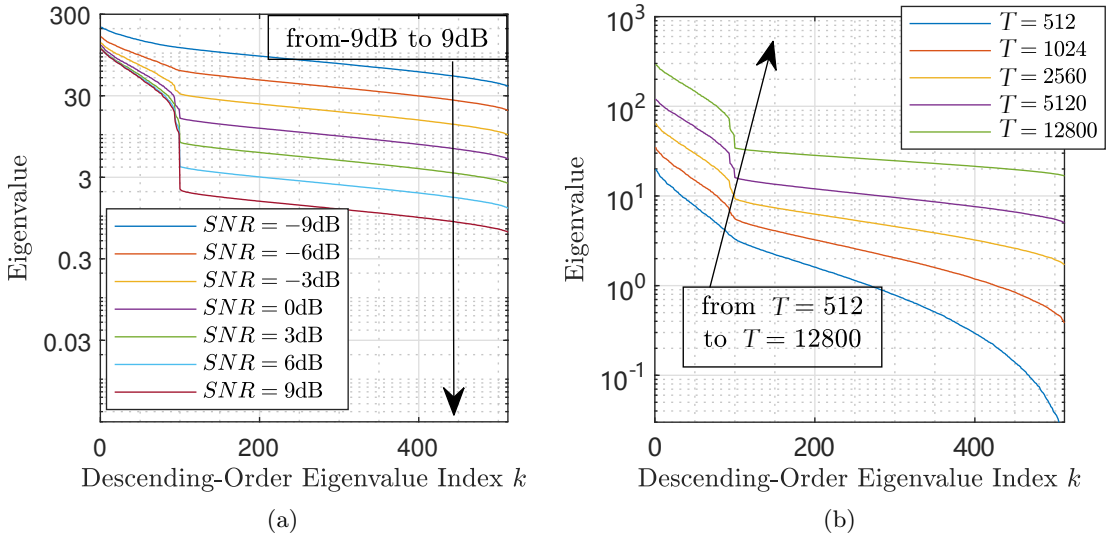


Figure 3.4: Descending-ordered eigenvalue profiles **(a)** with varying SNR from -9dB to 9dB and $T = 5120$; **(b)** with varying T and $SNR = 0$dB.

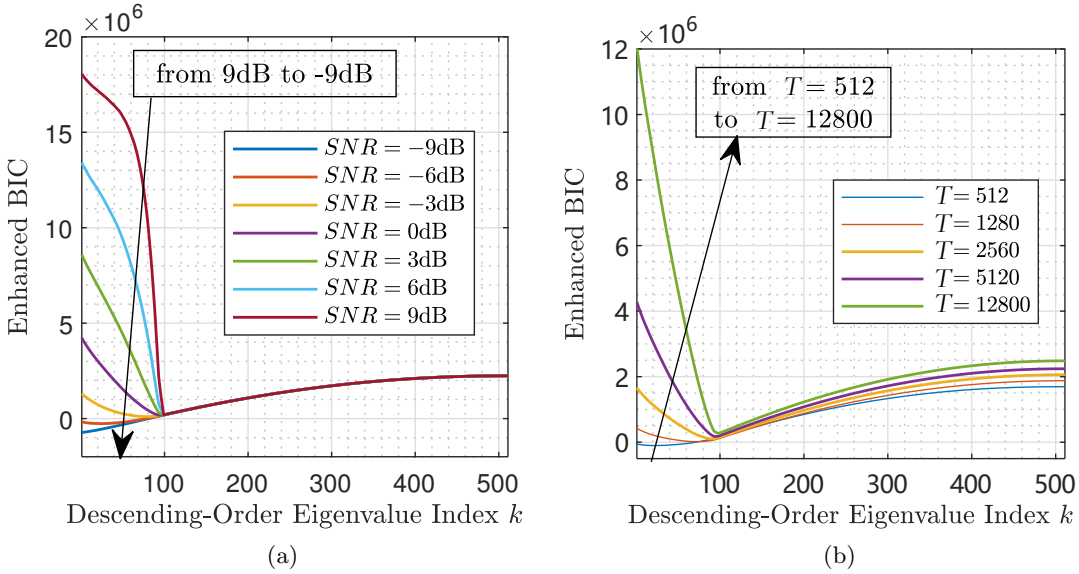


Figure 3.5: Enhanced BIC metric against sparsity values k **(a)** with varying SNR from -9dB to 9dB and $T = 5120$; **(b)** with varying T and $SNR = 0$dB.

integer. Hence, with negligible overestimation probability, it is proposed to estimate the active channel number l^o from the detected sparsity of spectrum by

$$l^o := \left\lceil \frac{k^o \cdot C}{N} \right\rceil . \tag{3.19}$$

This ceiling operation in active channel number detection effectively enhances the its robustness against low SNR and low values of T in which case the spectrum sparsity are possibly underestimated, by amending the underestimated spectrum sparsity $\delta(\boldsymbol{s}_f) - N/C < k^o < \delta(\boldsymbol{s}_f)$ as the case where the number of active channel is correctly detected. With the estimated active channel number l^o, one can get the compensated version of spectrum sparsity as

$$k^c := \frac{N \cdot l^o}{C} = \frac{N \cdot \left\lceil \frac{k^o \cdot C}{N} \right\rceil}{C}. \tag{3.20}$$

3.4 Low-Complexity CSS Scheme with Estimated Active Channel Number

In Section 3.3, it has been presented that one can apply a Bayesian-information-criterion-based approach to get valid sparsity estimates by detecting the 'best' dimension in the measurements' eigen-space that belongs to the signal components. This separation between the signal and noise subspace of the measurements indicates that one may remove the noise components and formulate a 'denoised' version of measurements for the benefit of performances of the following CS recovery procedures and the incumbent signal detection. Hereby, firstly, in Section 3.4.1, the noise reduction in the compressed measurements based on the valid sparsity estimation is presented, which effectively results in a new CS linear model with lower dimensions with lower recovery complexity expected. Next, in order to perform low-complexity sparse recovery based on the proposed noise-and-dimension-reduced model, greedy pursuit routines are investigated and empirically evaluated in Section 3.4.2. Then, in Section 3.4.3, a most interested greedy pursuit algorithm is chosen, which is further adapted to the newly proposed CS model to solve the joint-sparse spectrum signal in question. Finally, some useful conclusions on the convergence and complexity of the proposed recovery procedure are shown in Section 3.4.4.

3.4.1 Noise and Dimension Reduction based on Subspace Decomposition

Consider the sparse recovery based on the CS linear models. The introduction of the MMV model (3.9) can potentially improve the detection performance as the model includes multiple samples of random measurement vectors compared to SMV case (3.5). From the evaluations in Section 3.3.3, the proposed channel sparsity estimation may require a large number of measurement vectors T to achieve satisfactory performance. To enable quick spectrum sensing updates and rapid spectrum access, one could use a possibly smaller number of measurement vectors, i.e. $T^{'} \leq T$ to form the following MMV model

$$\boldsymbol{Y}^{'} = \boldsymbol{A}\boldsymbol{S}_{f}^{'} + \boldsymbol{B}^{'} = \boldsymbol{Y}_{s}^{'} + \boldsymbol{B}^{'}, \tag{3.21}$$

where $\boldsymbol{Y}^{'} = \left[\boldsymbol{y}^{(1)}\, \boldsymbol{y}^{(2)}\, \cdots\, \boldsymbol{y}^{(T^{'})}\right]$, $\boldsymbol{S}_{f}^{'} = \left[\boldsymbol{s_f}^{(1)}\, \boldsymbol{s_f}^{(2)}\, \cdots\, \boldsymbol{s_f}^{(T^{'})}\right]$, and $\boldsymbol{B}^{'} = \left[\boldsymbol{b}^{(1)}\, \boldsymbol{b}^{(2)}\, \cdots\, \boldsymbol{b}^{(T^{'})}\right]$. The sample acquisition period for (3.21) is then $T^{'}/F_s$.

However, noting that the sparse recovery's complexity via an MMV model grows linearly with the number of measurement vectors $T^{'}$ [16], sparse recovery via an MMV model (21) with a large $T^{'}$ becomes impractical. Besides that the subspace-aided approach in Section 3.3 provides the estimate of spectrum sparsity, it is hereby noted that by a similar subspace decomposition operation one can also reduce a great portion of noise from the measurements and also the dimension of CS recovery problem. Similar to (3.13), one has the estimated autocorrelation of $\boldsymbol{Y}^{'}$, which writes $\hat{\boldsymbol{R}}_{y}^{'} := \frac{1}{T^{'}} \sum_{p=1}^{T^{'}} \boldsymbol{y}^{(p)} \left(\boldsymbol{y}^{(p)}\right)^{H} = \frac{1}{T^{'}} \boldsymbol{Y}^{'} \boldsymbol{Y}^{'H}$, and perform eigendecomposition of $\hat{\boldsymbol{R}}_{y}^{'}$. With the compensated sparsity estimate k^c, one can separate the subspace of $\hat{\boldsymbol{R}}_{y}^{'}$ into signal and noise subspace

$$\begin{aligned} \hat{\boldsymbol{R}}_{y}^{'} =& \sum_{i=1}^{k^c} \hat{\lambda}_{i}^{'} \hat{\boldsymbol{v}}_{i}^{'} \hat{\boldsymbol{v}}_{i}^{'H} + \sum_{i=k^c+1}^{\operatorname{rank} \hat{\boldsymbol{R}}_{y}^{'}} \hat{\lambda}_{i}^{'} \hat{\boldsymbol{v}}_{i}^{'} \hat{\boldsymbol{v}}_{i}^{'H} \\ =& \boldsymbol{V}_s \boldsymbol{\Lambda}_s \boldsymbol{V}_s^{H} + \boldsymbol{V}_n \boldsymbol{\Lambda}_n \boldsymbol{V}_n^{H}, \end{aligned} \tag{3.22}$$

where $\hat{\lambda}_i^{'}$'s are descending-ordered eigenvalues and $\hat{\boldsymbol{v}}_i^{'}$ are corresponding eigenvectors, for $1 \leq i \leq \text{rank } \hat{\boldsymbol{R}}_y^{'}$; $\boldsymbol{\Lambda}_s = \text{diag}(\hat{\lambda}_1^{'}, \hat{\lambda}_2^{'}, \cdots, \hat{\lambda}_{k^c}^{'})$ and $\boldsymbol{\Lambda}_n = \text{diag}(\hat{\lambda}_{k^c+1}^{'}, \hat{\lambda}_{k^c+2}^{'}, \cdots, \hat{\lambda}_{\text{rank } \hat{\boldsymbol{R}}_y^{'}}^{'})$. The components in noise subspace can be easily identified and removed after k^c is obtained. In order to preserve the estimated dimension of signal subspace, i.e. k^c, it is required that rank $\hat{\boldsymbol{R}}_y^{'} \geq k^c$. On the other hand, from the dimension of $\hat{\boldsymbol{R}}_y^{'}$ itself, its rank satisfies rank $\hat{\boldsymbol{R}}_y^{'} = \min\left\{M, T^{'}\right\}$. Hence, it is necessary that the number of measurement vectors for sparse recovery meet $T^{'} \geq k^c$.

Then, the support of spectrum is proposed to be directly determined by the components in signal subspace $\boldsymbol{V}_s \boldsymbol{\Lambda}_s \boldsymbol{V}_s^H$, that is

$$\boldsymbol{V}_s (\boldsymbol{\Lambda}_s)^{\frac{1}{2}} = \boldsymbol{A}\boldsymbol{S}_v + \boldsymbol{B}_v, \tag{3.23}$$

where $(\boldsymbol{S}_v)_{N \times k^c}$ is the signal to be recovered by CS algorithm, and $\boldsymbol{B}_v$ denotes the perturbing term due to noise residual in the signal subspace. It is noted that the dimension of the joint-block-sparse matrix $\boldsymbol{S}_v$ in (3.23) instead of $\boldsymbol{S}_f^{'}$ (21) can be greatly reduced from $N \times T^{'}$ to $N \times k^c$. The relationship between the sparse signal of the two MMV models, $\boldsymbol{S}_v$ and $\boldsymbol{S}_f^{'}$ in (3.23) and (21) respectively, is expressed by

$$\text{supp}(\boldsymbol{S}_v) \subseteq \text{supp}(\boldsymbol{S}_f^{'}) = \text{supp}(\boldsymbol{S}_f), \tag{3.24}$$

where the equality holds when the spectrum sparsity is correctly estimated, i.e. $k^c = \delta(\boldsymbol{s}_f)$. With underestimated spectrum sparsity when SNR and T are small, the model (3.23) can still be used with possibly worse detection performance as the support of $\boldsymbol{S}_v$ is a subset of $\boldsymbol{S}_f$. The rationale of (3.24) is provided in Proposition 2 and Corollary 1 below[1].

Proposition 2. *For arbitrary matrices $\boldsymbol{M} \in \mathbb{C}^{a \times b}$ and $\boldsymbol{N} \in \mathbb{C}^{a \times c}$, if $\boldsymbol{M}\boldsymbol{M}^H = \boldsymbol{N}\boldsymbol{N}^H$, then span($\boldsymbol{M}$) = span($\boldsymbol{N}$).*

[1]The support of a joint-sparse matrix supp$(\cdot)$ is defined as the set of the indices of the rows whose 0-norm are non-zero.

Corollary 1. *If the MMV model (21) is transformed to (3.23) by zeroing the noise subspace in (3.22), and the dimensionality of signal subspace is not overestimated, i.e.* $k^c \leq \delta(\boldsymbol{s}_f)$*, then (3.24) is met. Equality in (3.24) is guaranteed if* $k^c = \delta(\boldsymbol{s}_f)$*.*

Proofs for Proposition 2 and Corollary 1 are given in the Appendix.

3.4.2 Evaluations on Greedy Algorithms

Here, four algorithms frequently referred to in literature are considered, namely Orthogonal Matching Pursuit (OMP), Compressive Sampling Matching Pursuit (CoSaMP) [147], Subspace Pursuit (SP) [148], and Hard Thresholding Pursuit (HTP) [149], all of which also require estimated sparsity as input. For simplicity, the basic SMV model (3.5) is examined here to compare these algorithms, where k^c is a natural choice of input sparsity estimate of $\boldsymbol{s}_f$. It is noted that the complexity of these algorithms is dominated by the correlation multiplication step with the complexity of $\mathcal{O}(NM)$, and the least-square estimation step as it refers to the computation of the pseudo-inverse of a $M \times k^c$ column-full-rank matrix - an $\mathcal{O}\left((k^c)^2 M\right)$ operation[2]. Without the knowledge of the signal sparsity, a safe practice is inputting an overestimated or the maximum allowed value as estimated sparsity in order to reduce the residual as much as possible. The difference between OMP and these three algorithms is that the OMP cumulatively detects the support in the desired signal, while CoSaMP, SP and HTP update a certain number of supports batch by batch until a satisfactory support is detected upon convergence. For OMP, the maximum allowed input of sparsity estimate is M with which a unique solution is marginally guaranteed in the least-square estimation step. The complexity reduction due to a valid estimated sparsity is apparent - the explicit number of iterations is reduced from M to k^c.

In the other three algorithms, in each iteration, the pruning operation is essential to update the support of the original sparse signal $\boldsymbol{s}_f$ where the elements of $\boldsymbol{s}_f$ corresponding to this support set are most likely to be non-zero. Specifically, CoSaMP has the pruning

[2]In this chapter, the pseudo-inverse of a $a \times b$ matrix ($b < a$) is considered to be implemented by the SVD-based algorithm, which gives the complexity of $\mathcal{O}(ab^2)$.

procedure where k^c most prominent supports out of $3k^c$ remain in each iteration. This pruning procedure dictates the maximum allowed input $k^c = M/3$, which implies sparsity can not be too much overestimated. Similarly, SP proposes to prune k^c out of $2k^c$ supports and then recalculates the least-squares in each iteration. In terms of HTP, the pruning process selects the k^c most relevant supports out of M. The complexity reduction comes from two aspects - one is the reduced dimensionality of the least-square estimation problem in each iteration from $\mathcal{O}(M^3)$ to $\mathcal{O}\left((k^c)^2 M\right)$, and another is the reduced average number of iterations as a valid estimate of sparsity aids to find the correct subspace dimension of the signal to be recovered.

In addition to the benefits of reduced complexity, the estimated sparsity information also aids to reliably recover the signal and find the correct signal support. Informally, for general greedy algorithms, inputting an overestimated signal's sparsity results in excessive free dimensions in the least-squares step, hence it leads to more inclusion of falsely-detected non-zero entries caused by noise perturbations in inaccurate measurements.

For direct empirical results, experiments are conducted for the SMV spectrum recovery problem (3.5) using the four interested greedy algorithms in order to compare their performance on recovery fidelity and running time. To provide fairness, the same stopping criteria are used for CoSaMP, SP and HTP - that the detected support is the same in two consequent iterations, or the algorithm reaches the maximum number of iterations of 100. These improvements on recovery fidelity illustrated in Fig. 3.6 is evaluated by normalized minimum-squared error (NMSE), defined by

$$NMSE = \frac{||\boldsymbol{s}_f^r - \boldsymbol{s}_f||_2^2}{||\boldsymbol{s}_f||_2^2}. \tag{3.25}$$

In Fig. 3.6 it is observed that all four algorithms shows the best recovery fidelity near accurate estimate of sparsity $k^c = \delta(\boldsymbol{s}_f) = 64$. For HTP, with the input sparsity estimate approaching M, the pruning is less effective and the recovery performance degrades

with M more severely compared with OMP. It worth noting that when inputting the sparsity estimate as M, the HTP algorithm degenerates to a single correlation step (which has only 2 iterations) and the recovery performance is the worst in terms of NMSE. However, with a valid sparsity estimate, the HTP algorithm offers the lowest NMSE, which is consistent with both the theoretical superiority [149, Th. 3.8] over the other three algorithms. For the time complexity of these algorithms, as shown in Fig. 3.7 (a) and (b), one can see that CoSaMP is sensitive to the input sparsity. Specifically, the detected support tends not to converge with overestimated input sparsity (reaching maximum iteration number of 100). Additional stopping criteria, which terminates the iteration when the residual is less than a predefined tolerance, may be desired to stop the algorithm before reaching the maximum number of iteration. However, the tolerance of the residual is related to the noise level, which requires prior knowledge of the signal. On the other hand, HTP still converges with small numbers of iterations. The good convergence performance of HTP and SP leads to the considerate less recovery time of both when compared with CoSaMP. Especially, OMP does not require termination upon convergence by running an explicit number of iterations which equals k^c, and the running time also largely exceeds that of HTP and SP.

3.4.3 HTP-Based Blind Block Support Detection

From both the theoretical guarantees and the empirical results above and in [149], it is found the HTP algorithm interesting which has superiority in both recovery fidelity and convergence with inaccurate sparsity input. Consequently, it is proposed that one can exploit the basic SMV form of HTP and revise it for solving the joint-block-sparse MMV model (3.23), namely joint-block-sparse hard thresholding pursuit (JB-HTP). Having obtained the estimated active channel number l^o from subspace-decomposition-based spectrum sparsity estimation, it is proposed to use l^o as the input to the proposed algorithm. The routines of the proposed algorithm are illustrated in Algorithm 1. To clarify the notations in the step 4, $(\cdot)_{\mathrm{BLK},\Omega^{(i+1)}}$ stands for the sub-matrix whose rows/columns are indexed by the block entries in the set $\Omega^{(i+1)}$, and $(\cdot)^\dagger$ denotes pseudo-inverse. In

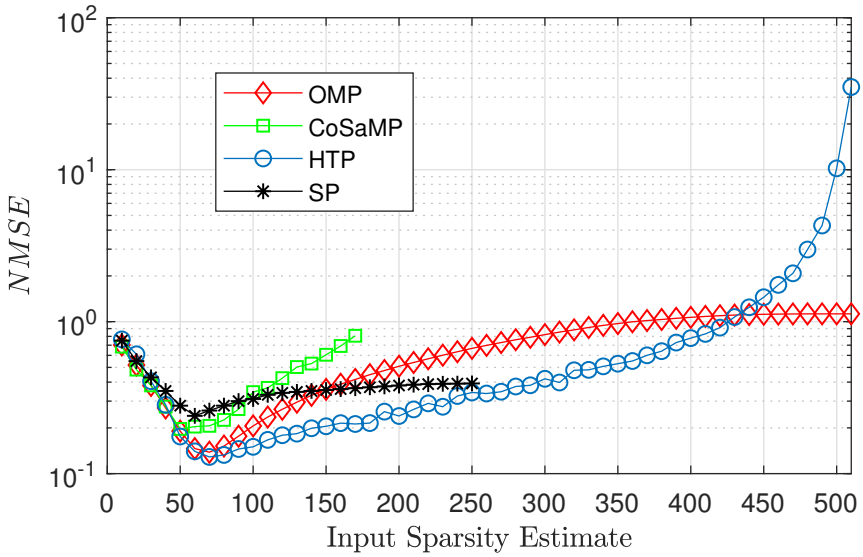


Figure 3.6: NMSEs of four greedy algorithms on the SMV model (3.5) against different input sparsity estimates k^c. Parameters are set as $N = 1280$, $M = 512$, $SNR = 6$dB and $\delta(\boldsymbol{s}_f) = 64$.

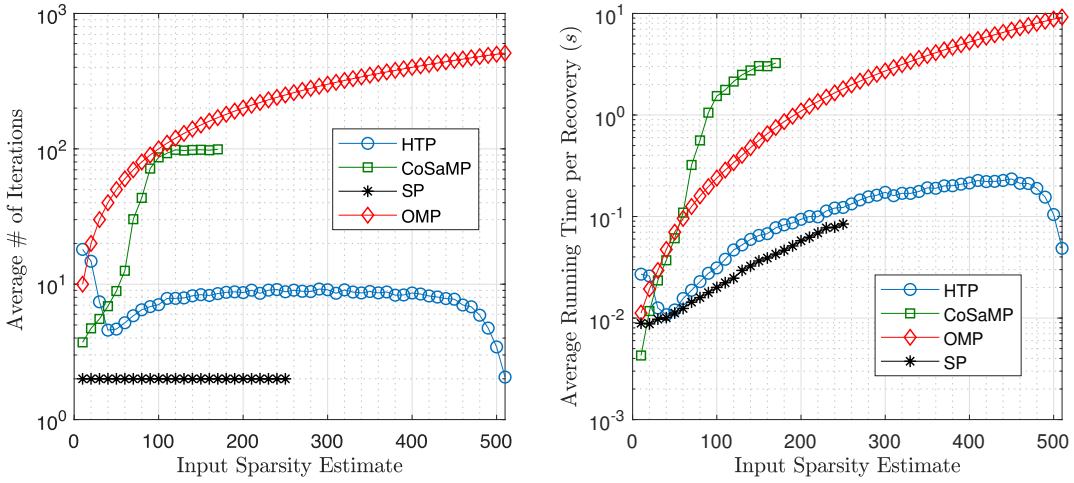


Figure 3.7: **(a)** Average number of iterations of CoSaMP, SP and HTP on the SMV model (3.5) against different input sparsity estimates k^c (maximum number of iteration is 100); **(b)** Average running time per recovery of four greedy algorithms on the SMV model (3.5) against different input sparsity estimates k^c.

the step 3 in the proposed algorithm, the Frobenius-norm is calculated for each block,

which writes

$$\left|\left|\left(\boldsymbol{S}_v^{(i+1)} + (\boldsymbol{A})^H \left[\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}} - \boldsymbol{A}\boldsymbol{S}_v^{(i+1)}\right]\right)_{\mathrm{BLK},\{c\}}\right|\right|_F \tag{3.26}$$

for $c = 0, 1, \cdots, C-1$. in comparison with the original HTP routine for SMV that employs 2-norm [149].

Algorithm 1 Joint-Block-Sparse Hard Thresholding Pursuit (JB-HTP)

Input: $\boldsymbol{\Phi}$, $\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}$, l^o, $maxIter$.
Output: $\boldsymbol{S}_v^r$, Ω.

1: Initialize $\boldsymbol{S}_v^{(i+1)} \leftarrow \boldsymbol{0}$, $\Omega^{(0)} \leftarrow \varnothing$, $\Omega_{\mathrm{old}} \leftarrow \varnothing$, $i \leftarrow 0$.
2: **while** $\left(\Omega^{(i)} \neq \Omega_{\mathrm{old}} \text{ and } i < maxIter\right)$ or $i = 0$ **do**
3: $\Omega^{(i+1)} \leftarrow \left\{\text{indices of } l^o \text{ largest } ||\cdot||_F \text{ block entries of } \boldsymbol{S}_v^{(i)} + (\boldsymbol{A})^H \left[\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}} - \boldsymbol{A}\boldsymbol{S}_v^{(i)}\right]\right\}$;
4: $\Omega_{\mathrm{old}} \leftarrow \Omega^{(i)}$;
5: $i \leftarrow i + 1$;
6: $\left(\boldsymbol{S}_v^{(i)}\right)_{\mathrm{BLK},\overline{\Omega^{(i)}}} \leftarrow \boldsymbol{0}$;
7: $\left(\boldsymbol{S}_v^{(i)}\right)_{\mathrm{BLK},\Omega^{(i)}} \leftarrow \left[(\boldsymbol{A})_{\mathrm{BLK},\Omega^{(i)}}\right]^{\dagger} \boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}$;
8: **end while**
9: $\Omega \leftarrow \Omega^{(i)}$;
10: $\boldsymbol{S}_v^r \leftarrow \boldsymbol{S}_v^{(i)}$;

Having finished the recovery of the joint-block-sparse matrix $\boldsymbol{S}_v$, one can perform ED based on the recovered matrix $\boldsymbol{S}_v^r$ to determine the channel occupancy. However, ED refers to setting an optimal threshold in the sense of detection probability, which requires extra knowledge of statistics of $\boldsymbol{S}_v$. Due to the nature of greedy pursuit, it is proposed to utilize the byproduct of the recovery algorithm - the detected block support Ω as the spectrum occupancy decision, which forms an HD scheme.

If the reconstructed spectrum $\left(\boldsymbol{S}_f^{'\,r}\right)_{N\times T'}$ analogous to the original joint-block-sparse spectrum $\boldsymbol{S}_f^{'}$ in the MMV model (21) is desired, it can be obtained by an additional single

projection step

$$\begin{aligned}\left(\boldsymbol{S}_f^{'\,r}\right)_{\mathrm{BLK},\overline{\Omega}} &\leftarrow \mathbf{0}\\ \left(\boldsymbol{S}_f^{'\,r}\right)_{\mathrm{BLK},\Omega} &\leftarrow \left[(\boldsymbol{A})_{\mathrm{BLK},\Omega}\right]^{\dagger}\boldsymbol{Y}_s^{'}.\end{aligned} \tag{3.27}$$

To summarize this section, the full procedures of the low-complexity blind CSS with spectrum sparsity estimation proposed in Section 3.3 and 3.4 are given in Alg. 2 and a complete overview of the proposed CSS system architecture is given in Fig. 3.1.

Algorithm 2 Low-Complexity Blind CSS with Spectrum Sparsity Estimation

Input: $M, N, T, \boldsymbol{\Phi}, \boldsymbol{Y}, C, maxIter$.
Output: Ω

1: Construct $\hat{\boldsymbol{R}}_y$ from T snapshots and perform EVD to obtain $\hat{\lambda}_1, \cdots, \hat{\lambda}_M$ and $\hat{\boldsymbol{v}}_1, \cdots, \hat{\boldsymbol{v}}_M$ as in (3.13);
2: Calculate BICe for $k = 1, \cdots M$ and find k^o corresponding the minimum BICe as in (3.18);
3: Compensate for underestimation and obtain l^o and k^c as in (3.19) and (3.20) respectively;
4: Construct $\boldsymbol{V}_s$ and $\boldsymbol{\Lambda}_s$ as in (3.22);
5: Recover the spectrum and detect incumbent blocks by $\{\boldsymbol{S}_v^r, \Omega\} \leftarrow$ JB-HTP $\left(\boldsymbol{\Phi}, \boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}, l^o, maxIter\right)$;

3.4.4 Theoretical Guarantees and Time Complexity of JB-HTP

Here, two conclusions of JB-HTP of interest are exhibited - firstly the exact recovery and convergence guarantee with accurate measurements of joint-block-sparse signals, and secondly the guarantee of the order of iteration times.

Proposition 3. *If the $3k^c$th-order restricted isometry constant[3] of the sensing matrix $\boldsymbol{A} := \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1} \in \mathbb{C}^{M\times N}$ satisfies $\sigma_{\boldsymbol{\Phi},3k^c} < 1/\sqrt{3}$, for any joint-block-sparse matrix $\boldsymbol{S}_v \in \mathbb{C}^{N\times k^c}$ with row sparsity of k^c, the solution $\boldsymbol{S}_v^{(i)}$ in the ith iteration of JB-HTP based on the MMV model (3.23) converges exactly to $\boldsymbol{S}_v$.*

[3]Defined in Lemma 2 in Appendix A.

See Appendix A for proof of Proposition 3.

Proposition 4. *If the $3k^c$th-order restricted isometry constant of the sensing matrix $\boldsymbol{A} := \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1} \in \mathbb{C}^{M\times N}$ satisfies $\sigma_{\boldsymbol{\Phi},3k^c} < 1/\sqrt{3}$, for any joint-block-sparse matrix $\boldsymbol{S}_v \in \mathbb{C}^{N\times k^c}$ with row sparsity of k^c and block sparsity of l, the average number of iterations required by the JB-HTP algorithm based on the MMV model (3.23) is at most $\mathcal{O}(\log l) + \mathcal{O}(1)$.*

See Appendix B for proof of Proposition 4.

Theoretically, each iteration of the proposed JB-HTP algorithm for MMV model (3.23) has complexity of $\mathcal{O}(k^cNM)$ from the matrix multiplication in the correlation step (line 3) and $\mathcal{O}\left((k^c)^2M\right)$ from pseudo-inverse in the least-square step (line 4). Consequently, each iteration's complexity writes $\mathcal{O}(k^cNM) + \mathcal{O}\left((k^c)^2M\right) = \mathcal{O}(k^cNM)$ as $k^c << N$ always holds. Hence, with Proposition 4, the total complexity of JB-HTP is expressed by $\mathcal{O}(k^cNM\log l)$.

3.5 Numerical Simulations

In simulations, time-domain OFDM signals $\boldsymbol{s}_o$ and $\boldsymbol{s}_t$ are generated as in (3.2) and (3.8) to provide a multiband signal model with sparse channel occupancy. Constant parameters are set as $B = 8$ (MHz), $C = 40$, $N = 1280$ and $CB/F_s = 7680$, and each row of $\boldsymbol{\Phi}$ is independent and generated by normalized Gaussian random vector $\mathcal{CN}(\mathbf{0}, \frac{1}{N}\mathbf{I})$. For simplicity of discussion, The wireless channel is considered to have unity gain and no multipath effect, i.e. $L_c = 1$ and $K_c = 0$ for $c = 1, 2, \cdots, C$ as in (3.4). Source symbols for the cth channel $\boldsymbol{\gamma}_c$ are set independent Gaussian, $\boldsymbol{\gamma}_c \sim \mathcal{CN}(\mathbf{0}, \sigma_\gamma^2\mathbf{I})$, if the cth channel is active or otherwise 0's in order to model a general multiband signal of which the spectrum coefficients in the active channels are Gaussian random. Additive white Gaussian noise model $\boldsymbol{b} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ is adopted and the received power, i.e. the variance of all γ_c's of all occupied channels are identical. Unless specified elsewhere, the default settings of parameters in the simulation are as follows. Actual active channel

number is set as $l = 3$, making $\delta(\boldsymbol{s}_f) = 96$ with central frequencies of incumbent channels set by $[f_1\ f_2\ f_3] = [76\ 164\ 244]$ (MHz), and the number of elements in each measurement vector is $M = 512$. The number of measurement vectors for sparse recovery in (21) is $T^{'} = M = 512$. The SNR, as formally defined in (3.6), can be revealed to be controlled by the ratio of σ_γ^2 and the noise variance σ^2 in the simulation setting,

$$SNR = \frac{\mathrm{E}\left[||\boldsymbol{\Phi}\boldsymbol{s}_t||_2^2\right]}{(M\sigma^2)} = \frac{\frac{M}{N}\mathrm{E}\left[||\boldsymbol{s}_f||_2^2\right]}{(M\sigma^2)} \underset{(3.7)}{=} \frac{\frac{M}{N}\cdot\frac{F_s N}{CB}\cdot\left(\frac{lB}{F_s}\sigma_\gamma^2\right)}{(M\sigma^2)} = \frac{l\cdot\sigma_\gamma^2}{C\cdot\sigma^2}.$$

3.5.1 Performance Evaluation of the Proposed Spectrum Sparsity Estimation Scheme

Firstly, as shown in Fig. 3.8, it is noted that the proposed spectrum sparsity estimation scheme accurately detects the true sparsity $\delta(\boldsymbol{s}_f) = lN/C = 96$ when SNR and the number of measurement vectors T are sufficiently large. Moreover, possible determinations of the active channel number $l^o = 1, 2, 3$ based on k^o have been shown. As a result of that the spectrum sparsity k^o is never overestimated, the proposed ceiling operation in (3.19) is justified and it provides more robustness against low SNR and T while still accurately detects the active channel number in high SNR and T cases.

In Fig. 3.9, the minimum integer SNRs with which the detection marginally gives correct active channel number with varying sample size T are shown. The results have verified the expectation that the decreasing M or increasing l lead to reduced dimensionality of the noise subspace, hence deteriorated performance.

3.5.2 Performance Evaluation of the Proposed CSS Scheme

3.5.2.1 Benchmark scheme settings

In order to evaluate the performance of the proposed JB-HTP-based CSS scheme, it is to be compared with a few chosen benchmark schemes illustrated in Table 3-A. In scheme No. I, II, and III, three CS recovery algorithms, BP, OMP and HTP are chosen

Table 3-A: Settings of the Proposed CSS Solution and Benchmark Schemes

Scheme No.	**CS Algorithm**	**CS Model**	**Detection Method**	k^c **available?**	**Recovery Complexity**
I	BP [55]	SMV-(3.5)	ED	Not Applicable	$\mathcal{O}(N^3)$ [15]
II	OMP [55]	SMV-(3.5)	ED	No, using $k^c = \mathcal{O}(M)$	$\mathcal{O}(M^4)$
III	HTP [149]	SMV-(3.5)	ED	No, using $k^c = \mathcal{O}(M)$	$\mathcal{O}(M^3 \log k^c)$
IV	OMP	SMV-(3.5)	HD	Yes	$\mathcal{O}(k^c NM)$ [148]
V	HTP	SMV-(3.5)	HD	Yes	$\mathcal{O}(NM \log k^c)$ [149]
VI	M-BP [173]	MMV-(3.23)	ED	Yes	$\mathcal{O}(k^c N^3)$
VII	SOMP [174]	MMV-(3.23)	HD	Yes	$\mathcal{O}\left((k^c)^2 NM\right)$
VIII (the proposed scheme)	JB-HTP	MMV-(3.23)	HD	Yes	$\mathcal{O}(k^c NM \log l)$

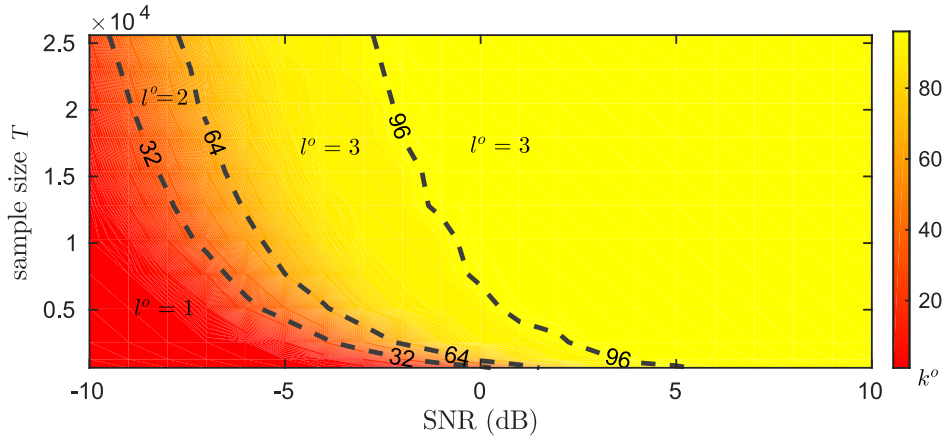


Figure 3.8: Estimated sparsity k^o and estimated number of active channels l^o against SNR and the number of snapshots for estimation T.

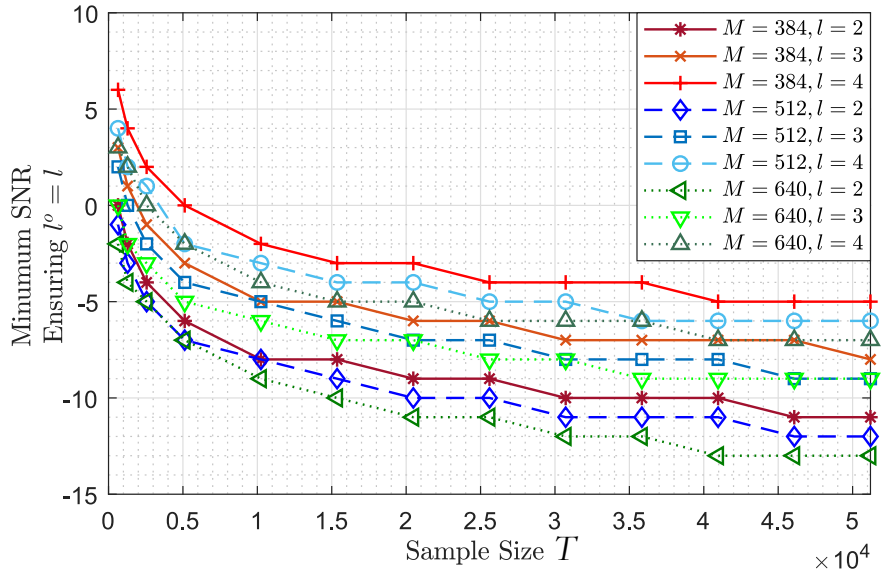


Figure 3.9: Minimum SNRs to ensure correct detection of active channel number $l^o = l$ against numbers of snapshots T, of varying numbers of elements each measurement vector M and l.

to solve the sparse recovery problem of the SMV model (3.5) without the knowledge of the spectrum sparsity where the ED scheme is used. To compare to SMV-ED schemes, using the active channel number estimation scheme proposed in Section 3.3, SMV-HD schemes with OMP and HTP recovery algorithms are examined in scheme No. IV and V. Apart from the proposed active channel number estimation scheme, in scheme No.

VI, VII, and VIII, the MMV model with dimension and noise reduction proposed in Section 3.4, i.e. (21), applies for sparse recovery. Particularly, in scheme No. VI, the MMV version of BP - multiple-BP (M-BP) [173] - is used to solve (3.9) where sparsity estimation is not applicable and hence ED has to be used. In scheme No. VI and VII, the MMV version of OMP, simultaneous-OMP (SOMP) [174], and the proposed JB-HTP are used for MMV model (21) where HD is used because a spectrum sparsity estimation is required by the formulation of the proposed model (21).

It should be noted that the recovery algorithms - OMP, HTP and SOMP in scheme No. IV, V and VII are not intended for block-sparse signals so that the block support (i.e. the active channel support) is not directly output by these recovery algorithms. In these schemes, an intuitive procedure to produce the active channel support is used here as in [54]. Having recovered the spectrum $\boldsymbol{s}_f^r = [s_f^r[0]\ s_f^r[1] \dots s_f^r[N-1]]^T$ from (3.5) or $(\boldsymbol{S}_v^r)_{N\times k^c}$ from (3.23), the channel energy is calculated by $\boldsymbol{p}_f = [p_f[0]\ p_f[1] \dots p_f[N-1]]^T$ whose entries are

$$p_f[c] = \sum_{i=cN/C}^{(c+1)N/C-1} |s_f^r[i]|^2, \tag{3.28}$$

for SMV model (3.5) or

$$p_f[c] = \left\| (\boldsymbol{S}_v^r)_{\mathrm{BLK},\{c\}} \right\|_F^2, \tag{3.29}$$

for MMV model (3.23), where $c = 0, 1, \cdots, C-1$. One can then directly determine the active channels by sorting $p_f[n]$'s and indicating the channels corresponding to the largest l^o entries, which effectively achieves hard detection, formally

$$\Omega = \{\omega_1, \cdots, \omega_{l^o} | p_f[\omega_1] \geq \cdots \geq p_f[\omega_{l^o}] \geq \cdots \geq p_f[\omega_C]\}. \tag{3.30}$$

3.5.2.2 Evaluation Metrics

For the purpose of spectrum sensing in DSA applications, the performance of detecting occupied and vacant channels is the natural choice of the performance metrics. In Monte-Carlo simulations, by definition, the probability of detection P_d and the probability of

false alarm P_f are presented by

$$P_d = \frac{\texttt{total \# correctly detected channels}}{\texttt{\# recovery trials} \times l},$$
$$P_f = \frac{\texttt{total \# falsely detected channels}}{\texttt{\# recovery trials} \times (C - l)}.$$

3.5.2.3 Simulation results

For the general effectiveness of the proposed scheme VII with JB-HTP, the detection performance with different numbers of samples per measurement vector M and block sparsity l of the spectrum is shown in Fig. 3.10. One can observe that the default setting ($M = 512, l = 3$) gives P_d close and approaching 1 ($P_d > 0.99$) and P_f close and approaching 0 ($P_f < 10^{-3}$) starting from the SNR as low as -20dB. In comparison with other combinations of M and l, the proposed scheme, like other CSS schemes, shows superior detection performance with larger M (i.e. larger compression ratio) and smaller l (i.e. smaller spectrum sparsity). It should be made clear that the estimated active channel number used here is accurate, that is $l^o = l$ being input to Algorithm 1.

Next, the detection performance of the SMV schemes is examined to exhibit the benefit of using the HD scheme brought by the proposed active channel number estimation. Receiver operating characteristics (ROCs) of HD scheme IV and V are given in Fig. 3.11 to compare with ROC curves of ED schemes - I, II, and III. Note that ROCs of HD schemes are scattered points as it is a hard decision scheme. Within the ED schemes, the BP algorithm outperforms the greedy algorithms in scheme II and III. Moreover, the advantage in detection performance of HD schemes is obvious as each ROC scatter lies in the upper-left region to its counterpart ROC curve of the ED scheme, and even the ED scheme with BP, under the same SNR. Specifically, the proposed HD scheme can achieve the same P_d level while dramatically reduces P_f by orders of magnitude. This advantage may be intuitively credited to the aid of the extra information of the active channel number estimation to both the recovery and decision-making procedures. It is

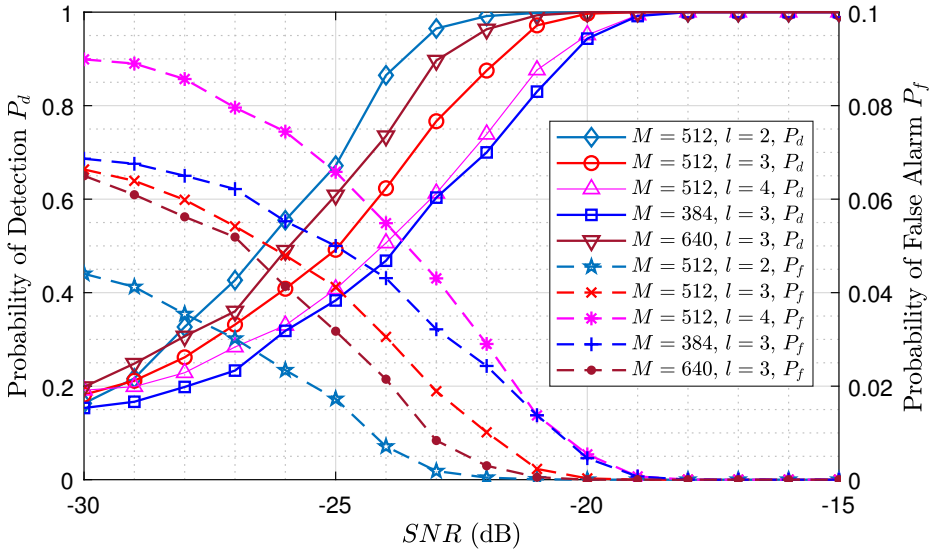


Figure 3.10: Detection performance of the proposed JB-HTP scheme on the MMV model (3.23) against SNR. It is assumed the active channel number are correctly estimated, $l^o = l$ in all shown cases.

also noted that the effect of the hard decision based on l^o intrinsically constrained P_f with P_d, that is $P_f = (l^o - lP_d)/(C - l)$. This implies that the proposed hard-decision scheme has constrained and small P_f when P_d is relatively large and l is only a fraction of C. The constraint for $l^o = l = 3$ is drawn in Fig. 3.11 where $P_f < 0.016$ is guaranteed when $P_d > 0.8$. This inherent constraint on P_f is desirable compared to the ED scheme, where setting a threshold with P_f constraints requires knowledge of the noise statistics.

Furthermore, one would like to account for the benefit of implementing the noise and dimension reduction (3.22) and using the MMV model (3.23) by comparing the detection performance of MMV-HD schemes - VII and VIII - with that of SMV-HD schemes - IV and V. In Fig. 3.12, it is clear that each MMV scheme - the proposed JB-HTP or SOMP - shows evidently higher robustness regarding detection performance against noise. Taking $P_d > 0.99$ and $P_f < 10^{-3}$ as reference levels, quantitatively, the proposed CSS scheme (VIII) with JB-HTP shows the superior robustness against the noise of as much as 20 dB to achieve equivalent detection performance in comparison to both SMV-HD schemes

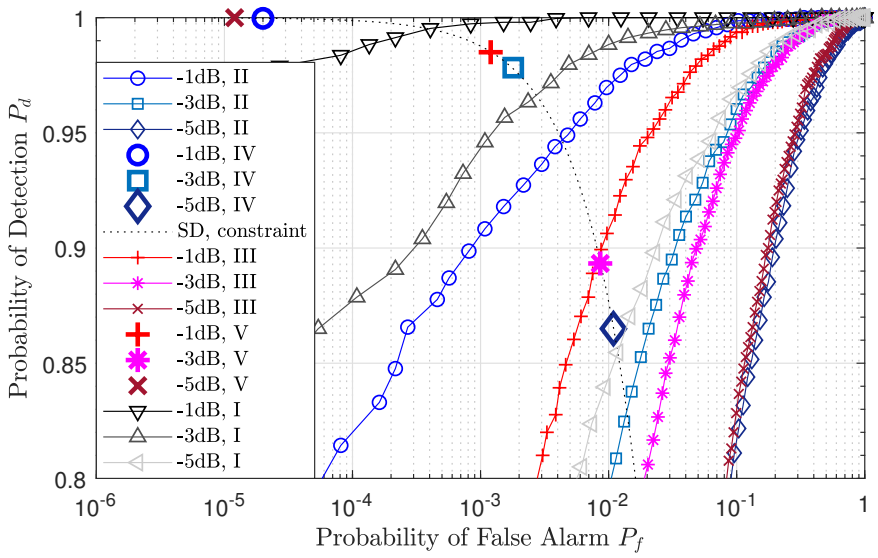


Figure 3.11: ROC curves of SMV-ED schemes and scatterplots of SMV-HD schemes in Table $3-A$ with SNRs from -1dB to -5dB and $l^o = l = 3$.

with OMP and HTP. On the other hand, Fig. 3.12 indicates that the use of the proposed JB-HTP algorithm based on the MMV model (3.23) proves to deliver better detection performance than using SOMP in scheme No. VII. Furthermore, results have also been obtained for the proposed scheme (VIII) with different numbers of measurement vectors, $T^{'} = 256$, 512, and 1024. As expected, a larger number of stacked measurement vectors in (21) leads to better separation of signal and noise components via eigendecomposition hence better detection performance after sparse recovery, but one should bear in mind that a large $T^{'}$ also implies longer sample acquisition time and consequently delays in obtaining spectrum sensing results.

In Fig. 3.13, the detection performance of MMV schemes VI, VII, and VIII are illustrated in the form of ROC scatterplots and curves. The results show that the proposed scheme VIII with JB-HTP outperforms scheme VII with SOMP and scheme VI with M-BP and ED in respect of detection performance, while scheme VII is only slightly better than the ED scheme VI. Additionally, it is also worth considering the cases when l^o is underestimated due to low SNR or the number of measurement vectors T, and the possible option of purposely overestimating l^o. With underestimated l^o,

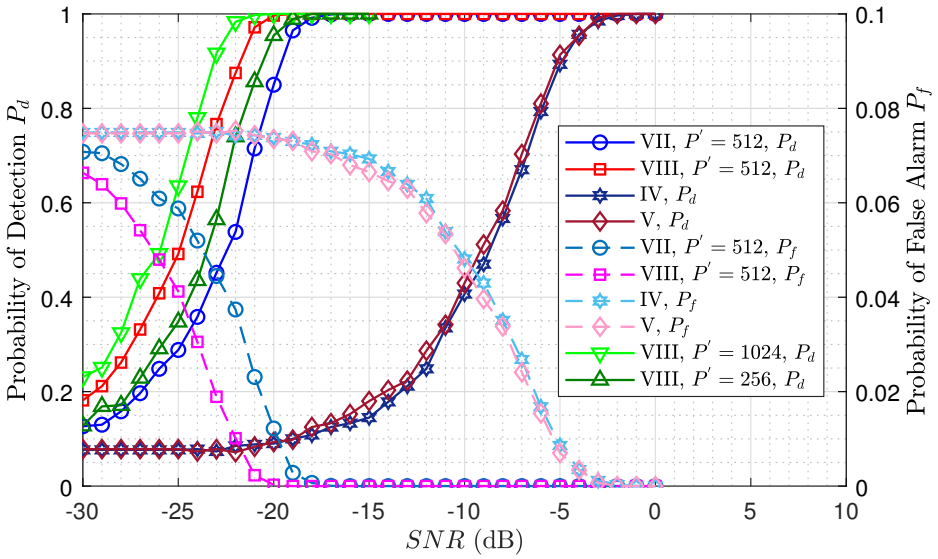


Figure 3.12: Detection performance against SNR of MMV-HD schemes compared with SMV-HD counterparts. It is assumed the active channel number are correctly estimated, $l^o = l$ in all shown cases.

any HD scheme, can only achieve relatively poor P_d upper-bounded by l^o/l, as a direct consequence of at least $l - l^o$ miss-detected active channel. However, one can deliberately compensate l^o by some integer where some levels of underestimation are prone to appear under certain SNRs. The price of such compensation method is that l^o is also likely to be overestimated. Results are also obtained for the proposed scheme VII to evaluate the detection performance with both underestimate ($l^o = l - 1 = 2$) and overcompensate ($l^o = l + 1 = 4$) cases and results are also given in Fig. 3.13. In the underestimate case, P_d's deterioration is obvious - it can only achieve no higher than 0.67 as expected. In the overcompensate case, the price is the increased P_f which is lower-bounded by $1/(C - l) = 0.027$, and P_d is somehow greater, compared to the case where $l^o = 3$ under the same SNRs.

Finally, the complexity of these benchmarked schemes I-VII and the proposed scheme VIII are investigated. The time complexity of each CSS scheme in Table 3-A has been accounted for in the order of atomic manipulations of complex addition and multiplication. In comparison within either SMV or MMV schemes, the HTP or the proposed

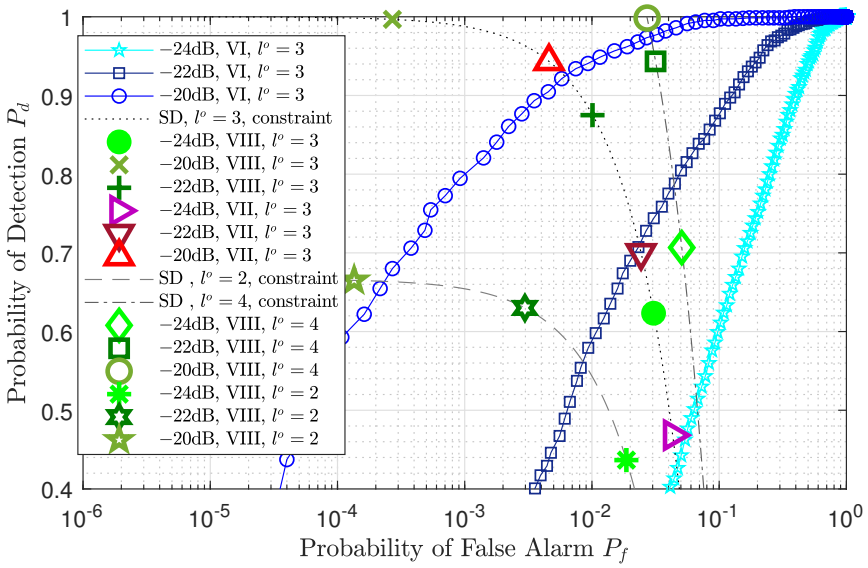


Figure 3.13: ROC curves of MMV-ED schemes and scatterplots of MMV-HD schemes in Table $3-A$ with SNRs from -20dB to -24dB and $l^o = l = 3$. ROC scatterplots of MMV-HD schemes with underestimated active channel number $l^o = l - 1 = 2$ and overcompensated active channel number $l^o = l + 1 = 3$ are also shown for comparison.

JB-HTP scheme has the smallest order of time complexity in the recovery phase. Note there is also an additional complexity of $\mathcal{O}(M^3)$ in scheme No. IV-VIII from the active channel number estimation due to the SVD procedure. In addition, the empirical results on the average time per recovery of the 8 schemes are shown in Fig. 3.14 with 3 different values of M. The time required for estimating the active channel number has also been accounted for schemes where applicable. As a general trend, the average time consumed per recovery is positively related to the value of M. More importantly, the empirical time complexities, with the account of estimation of l^o, of scheme No. VIII and VI are also the smallest among MMV and SMV schemes, respectively. By comparing the proposed MMV scheme VIII with the SMV counterparts III and VI, although the analytic complexity order and recovery time of scheme VIII is found slightly greater as shown in Fig. 3.14, it is reminded by the earlier experiments in Fig. 3.11 and 3.12 that the superiority of the scheme VIII in detection performance is tremendous.

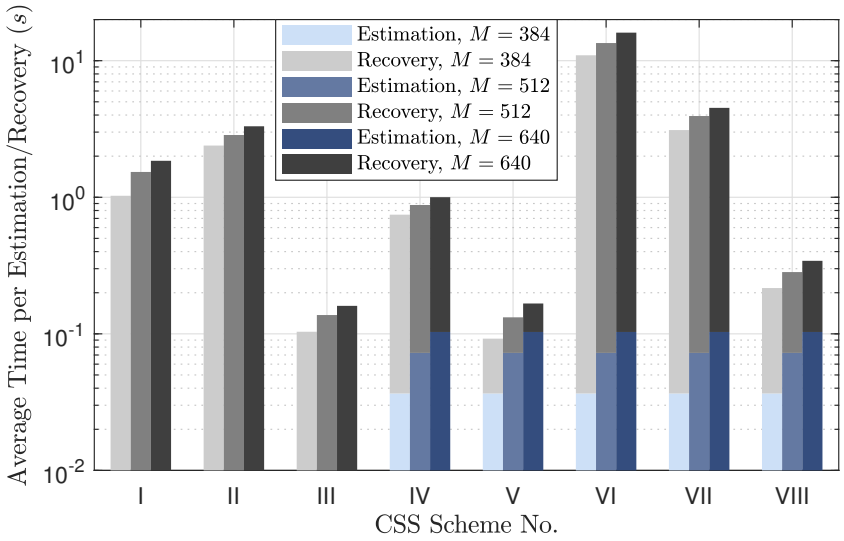


Figure 3.14: Average time consumed by the CS recovery schemes in Table 3-A together with the active channel number estimation scheme (if applicable).

3.6 Summary

This chapter presents a novel greedy-pursuit-based CSS scheme with the aid of subspace-decomposition-based spectrum sparsity estimation, where the spectrum sparsity is directly estimated from the output of the sub-Nyquist measurements without recovery operations. In the proposed scheme, an MMV model with noise and dimension reduction is introduced, and a generalized version of HTP for joint-block sparse signals is proposed as the recovery algorithm. The benefits of such a novel CSS scheme are summarized in three aspects. Firstly, the estimation of spectrum sparsity effectively alleviates the computational complexity of spectrum recovery with an explicit spectrum sparsity estimate compared to the cases where such information is absent. Secondly, the use of the proposed MMV model with noise and dimension reduction further enhances the detection performance (as much as 20dB seen in simulation) while effectively keeping the complexity low. Thirdly, the use of the spectrum sparsity estimate enables hard detection when determining the occupancy of wireless channels, which improves the detection performance from that of energy detection schemes, as well as removes the need for threshold adaption. Monte-Carlo simulations have evaluated the performance of the spectrum

sparsity estimation, shown the superior detection performance of the proposed scheme against multiple benchmarking schemes, and verified its low time complexity.

Chapter 4

Channel Energy Modeling and Threshold Adaption in Compressive Spectrum Sensing

Spectrum sensing is a proactive way in cognitive radio systems to achieve dynamic spectrum access, and compressive spectrum sensing (CSS) techniques alleviate the demand for high-speed sampling in wideband spectrum sensing. Most of the existing literature discusses Neyman-Pearson channel energy detection and threshold adaption schemes to achieve optimal detection performance in conventional non-compressive spectrum sensing scenarios. However, in the CSS, it is found that the channel energy statistics and optimal threshold not only depend on noise energy but also the compression ratio, the sparsity of spectrum, and the nature of recovery algorithms. To investigate the channel energy statistics of the recovered spectrum, a statistical model of channel energy for CSS is postulated and a learning algorithm based on Mixture Model and Expectation-Maximization techniques is proposed. Additionally, having verified the validity of the postulated model, a practical threshold adaption scheme for CSS is proposed aiming to keep constant false alarm rates in channel energy detection. In simulations, it is

shown that the postulated channel energy statistical models with parameters learned by the proposed learning algorithm fit well with empirical distributions for various channel models and recovery algorithms. Moreover, it is shown that the proposed threshold adaption scheme keeps the false alarm rate near the predefined constant, which in turn validates the postulated model. The rest of this chapter is organized as follows. Section 4.1 outlines the introduction and contributions of the work in this chapter. In Section 4.2, the signal model and the NP energy detection problem in CSS are illustrated. In Section 4.3, the postulated statistical model for recovered signal in CSS is presented and an algorithm for learning the parameters in the postulated statistical model is proposed. In Section 4.4, based on the results in Section 4.3, the threshold adaption scheme based on noise statistics estimation is proposed. Simulation results to prove the effectiveness of the postulated channel energy statistical model and the proposed learning algorithm, as well as the performance of the proposed threshold adaption scheme, is shown in Section 4.5.

4.1 Introduction

4.1.1 Related Work

statistical model Detection of occupied channels using spectrum reconstructed by CS recovery algorithms is a crucial procedure of CSS. Cyclostationary feature detection (CFD) and channel energy detection are two major types of detection techniques applied in CSS that are reported in the literature. CFD exploits the cyclic stationary property in modulated radio signals and it conducts detection in the spectrum of cyclic frequency and spectral frequency [103, 104]. Compared with energy detection, CFD has been shown to perform better in lower signal-to-noise ratio (SNR) conditions [69, 175]. However, it adds considerable computational complexity to receivers and CFD-based detection algorithms only work with specific and known types of modulation of signals. The simpler yet more commonly used method is Neyman-Pearson (NP) channel energy detection whose objective is to directly differentiate the present signal from noise in a certain channel's power

spectrum by setting a proper threshold [97]. A common problem of energy detectors is that statistics of the noise, for example, the variance, are a priori unknown in most cases, because the noise in the received signal depends on the receiver's noise figure and gain control, temperature, ambient radio interference, etc. Thus, the noise statistics need to be estimated to achieve optimal detection performance. As in most literature of conventional spectrum sensing, it is assumed the noise is Gaussian distributed in the NP energy detection method and the detection performance can be evaluated for various channel models [100, 160]. Under the assumption of Gaussian noise, it has been proposed in [100, 160, 176, 177] that noise variance estimation and threshold adaption can be achieved in an online fashion. Other discussions on threshold adaption schemes without prior assumptions on signal statistics are seen in the literature where supervision in the adaptation process is necessary [103, 178]. However, the supervised learning process requires training radio pilots of which the spectrum occupancy information is readily known to sensing nodes, which is unrealistic in real-world spectrum sensing applications.

In CSS, the desired spectrum for energy detection is not directly available from time-domain samples of analog-to-information converter (AIC) [71]. The spectrum of interest needs to be recovered from these sub-Nyquist-rate samples by sparse recovery algorithms. From both restricted isometry property theory of CS [179] and practice of wideband CSS [14–16, 59, 180], the sparse recovery of a noisy signal results in an inconsistent recovered spectrum compared to the true spectrum, and such inconsistency depends on spectrum sparsity, compression ratio and SNR of the sensed signal. From past experiments [15, 181], it is discovered that in CSS the energy statistics of the recovered spectrum differ from that of the true spectrum. In order to achieve optimal energy detection performance, a proper statistical model and unknown statistics in the recovered spectrum need to be learned.

4.1.2 Contributions

The main contribution of this chapter is that, to the best knowledge, it is the first work to address and model the statistics of the recovered signal in the energy detection problem of CSS. It is discovered that the channel energy statistics in CSS are fundamentally different from that in conventional non-compressive spectrum sensing. Specifically, to set NP energy detection hypotheses for recovered signals, it is postulated that the channel energy statistics model of the recovered spectrum still conforms to that of the original spectrum, however, parameters of the model for recovered signals are treated as unknown. Mixture Model (MM) and Expectation-Maximization (EM) techniques [182] have been commonly used to obtain maximum-likelihood estimates of the parameters given analytic distributions of statistics, and the specific use of Rayleigh-Gaussian MM has been seen in [183, 184] to learn the signal statistics in non-compressive spectrum sensing. In this chapter, the focus is on the channel energy in CSS and a customized EM-based algorithm for a chi-square-MM is proposed to learn the channel energy statistics of the recovered signal. Simulations have shown that the postulated statistical model for recovered spectrum is a reasonably good fit with parameters learned from the sample data set by the proposed algorithm.

An additional contribution is that, furthermore, a novel and practical threshold adaption scheme is proposed based on the newly-addressed statistical model to achieve the detection performance of constant false alarm rate (CFAR) for energy detection in CSS. In simulations, it is shown how differently the thresholds should be set in various settings of the CSS. Moreover, the results that the probability of false alarm can be kept near the predefined constant also validate the good fit of the postulated model of channel energy statistics and learning algorithm.

4.2 Signal Model and Problem Statement

4.2.1 Compressive Spectrum Sensing

Consider CSS by single sensing node in non-cooperative scenario. Denote the Nyquist time-domain signal as $\boldsymbol{s}_t = [s_t^{(1)}\ s_t^{(2)} \dots s_t^{(N)}]^T$ and its frequency-domain representation $\boldsymbol{s}_f = \boldsymbol{\mathcal{F}}\boldsymbol{s}_t = [s_f^{(1)}\ s_f^{(2)} \dots s_f^{(N)}]^T$ where $\boldsymbol{\mathcal{F}}_{N\times N}$ stands for N-point Discrete Fourier Transform (DFT) matrix. A signal sparse in frequency domain can be recovered based on M sub-Nyquist-rate measurements $\boldsymbol{y} = [y^{(1)}\ y^{(2)} \dots y^{(M)}]^T$ where $M < N$. The sub-Nyquist sampling can be expressed as a linear system

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{s}_t + \boldsymbol{b} = \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}\boldsymbol{s}_f + \boldsymbol{b}, \tag{4.1}$$

where $\boldsymbol{\Phi}_{M\times N}$ is the sampling matrix with structured random entries corresponding to AIC sampler settings, and $\boldsymbol{b} = [b^{(1)}\ b^{(2)} \dots b^{(M)}]$ is additive noise present in measurements.

The recovery of the sparse signal $\boldsymbol{s}_f$ can be achieved by solving the optimization problem, writing

$$\boldsymbol{s}_f^r = \arg\min_{\boldsymbol{s}_f} ||\boldsymbol{s}_f||_l + \frac{1}{2\lambda}||\boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}\boldsymbol{s}_f - \boldsymbol{y}||_2^2, \tag{4.2}$$

where the norm $0 < l \leq 1$ and λ accounts for the penalty function and $\boldsymbol{s}_f^r$ is the recovered version of the original signal. Specifically when $l = 1$, the optimization problem is convex and can be solved by Basis Pursuit Denoising algorithm. Furthermore, when $\lambda = 0$, which is usually adopted as penalization term is often a priori unknown, the recovery algorithm is called Basis Pursuit (BP). Besides, greedy algorithms are also an efficient category of sparse recovery solvers, among which the most commonly used is the Orthogonal Matching Pursuit (OMP) [55]. Sparse Bayesian Learning (SBL) has been recently proposed for solving the sparse recovery problem in a probabilistic setting. By introducing the Gaussian assumption on noise, specifically $p(\boldsymbol{y}|\boldsymbol{s}_f)$ being Gaussian and assuming parameterized Gaussian priors $\boldsymbol{s}_f \sim \mathcal{CN}(\boldsymbol{0}, \mathrm{diag}(\boldsymbol{\gamma}))$ that induces sparsity in the recovered signal, it aims to find maximum a posteriori probability of the hyperparameters

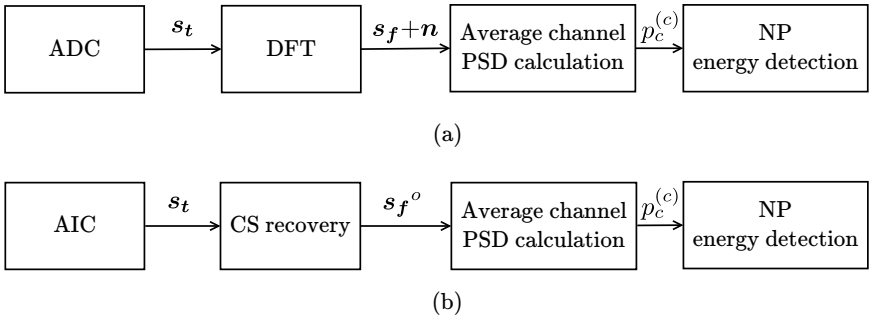


Figure 4.1: System architecture of **(a)** conventional non-compressive compressive sensing and **(b)** CSS.

$\boldsymbol{\gamma}$ [91, 94]:

$$\boldsymbol{s}_f^r = \arg\max_{\boldsymbol{\gamma}} p(\boldsymbol{\gamma}|\boldsymbol{y}) = \arg\max_{\boldsymbol{\gamma}} \int p(\boldsymbol{y}|\boldsymbol{s}_f) p(\boldsymbol{s}_f;\boldsymbol{\gamma}) \mathrm{d}\boldsymbol{s}_f. \tag{4.3}$$

4.2.2 Channel Energy Detection in Spectrum Sensing

Consider wideband spectrum sensing for multiband signals where a channel is the unit of frequency resource. To detect the channel occupancy, one can calculate the average of the power spectrum density (PSD) bins within each channel for NP energy detection as

$$p_c^{(c)} = \sum_{(c-1)R \le i < cR} \frac{1}{R} |x_f^{(i+1)}|^2, \tag{4.4}$$

for $c = 1, 2, \ldots, C$, where C is the total number of channels and R is the number of PSD bins in each channel. The system architecture diagrams of non-compressive spectrum sensing and CSS with channel energy detection are illustrated in Fig. 4.1 (a) and (b) respectively.

Next, the statistical model of channel energy is formulated for conventional non-compressive spectrum sensing, and the differences in the channel energy statistics for compressive spectrum sensing cases are described.

For non-compressive spectrum sensing, express the sensed spectrum as

$$\boldsymbol{x}_f = \boldsymbol{s}_f + \boldsymbol{n}_f = \boldsymbol{\mathcal{F}} \boldsymbol{x}_t = \boldsymbol{\mathcal{F}}(\boldsymbol{s}_t + \boldsymbol{n}_t), \tag{4.5}$$

where $\boldsymbol{n}_{t_{N\times 1}} = [\omega_t^{(1)}\ \omega_t^{(2)} \cdots \omega_t^{(N)}]^T$ denotes complex independent-and-identically-distributed (i.i.d.) additive white Gaussian noise (AWGN) as is widely adopted in the literature [15, 97, 100]. Denote $\boldsymbol{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. Note that the dimension of $\boldsymbol{x}_f$ here is $N = C \cdot R$. Consider a general multipath scenario, the expressions of $x_t^{(n)}$ of the null hypothesis $\mathcal{H}_0$ corresponding to absent radio activity and $\mathcal{H}_1$ corresponding to active radio activity, are written by

$$x_t^{(n)} = \begin{cases} w_t^{(n)} & \mathcal{H}_0 \\ \sqrt{K}e_t^{(n)} + \sqrt{1-K}e_t^{(n)}h_t^{(n)} + w_t^{(n)} & \mathcal{H}_1 \end{cases}$$

for $n = 1, 2, \cdots, N$, where a commonly-adopted model

$$\boldsymbol{h}_t = [h_t^{(1)}\ h_t^{(2)} \cdots h_t^{(N)}]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$$

is considered that characterizes the multipath effect of channel. K is the power ratio of line-of-sight versus multipath components. The vector $\boldsymbol{e}_t = [e_t^{(1)}\ e_t^{(2)} \cdots e_t^{(N)}]^T$ represents the deterministic samples in a time frame of the PUs' transmitted signal attenuated by channel gain, and it is defined that $\boldsymbol{e}_f = [e_f^{(1)}\ e_f^{(2)} \cdots e_f^{(N)}] := \boldsymbol{\mathcal{F}}\boldsymbol{e}_t$. After performing a linear transform (i.e. the DFT) on $\boldsymbol{s}_t$ and some rearrangements, one can reach the statistical model of $\boldsymbol{x}_f$

$$x_f^{(n)} = \begin{cases} w_f^{(n)} \sim \mathcal{CN}(0, N\sigma^2) & \mathcal{H}_0 \\ \sqrt{K}e_f^{(n)} + \sqrt{1-K}e_f^{(n)}h_f^{(n)} + w_f^{(n)} & \\ \quad \sim \mathcal{CN}(\sqrt{K}e_f^{(n)}, (1-K)|e_f^{(n)}|^2 + N\sigma^2) & \mathcal{H}_1 \end{cases} \tag{4.6}$$

where one can find

$$\boldsymbol{h}_f = [h_f^{(1)}\ h_f^{(2)} \cdots h_f^{(N)}]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$$

and denote

$$\boldsymbol{n}_f = [\omega_f^{(1)}\ \omega_f^{(2)} \cdots \omega_f^{(N)}]^T := \boldsymbol{\mathcal{F}}\boldsymbol{n}_t \sim \mathcal{CN}(\mathbf{0}, N\sigma^2\mathbf{I})$$

.

Moreover, from (4.4), the statistical model of average channel PSD level $p_c^{(c)}$ is found characterized by a central and a noncentral chi-square distribution, as a direct result of summing the squared Gaussian distributions of $\boldsymbol{x}_f$ in (4.6) with zero means and non-zero means respectively. Specifically, it writes

$$p_c^{(c)} = \begin{cases} \sum\limits_{(c-1)R \le i < cR} \frac{|w_f^{(i+1)}|^2}{R} = \frac{N\sigma^2}{2R} r_0,\ r_0 \sim \chi^2(2R) & \mathcal{H}_0 \\ \sum\limits_{(c-1)R \le i < cR} \frac{1}{R} |\sqrt{K} e_f^{(i+1)} + \\ \qquad \sqrt{1-K} e_f^{(i+1)} h^{(i+1)} + w_f^{(i+1)}|^2 & \mathcal{H}_1 \\ \qquad = \beta^{(c)} r_1,\ r_1 \sim \chi'^2(2R, \alpha^{(c)}) \end{cases} \tag{4.7}$$

for $c = 1, 2, \cdots, C$, where $\chi^2(k)$ and $\chi'^2(k, q)$ represents central chi-square distribution of degree of freedom (DoF) k and non-central chi-square distribution of DoF k and noncentrality parameter q respectively. In (4.7), $r_0 \sim \chi^2(2R)$ and $r_1 \sim \chi'^2(2R, \alpha^{(c)})$ are random variables. Parameters $\alpha^{(c)}$ and $\beta^{(c)}$ are expressed as

$$\alpha^{(c)} = \sum_{(c-1)R \le i < cR} \frac{K |e_f^{(i+1)}|^2}{(1-K)|e_f^{(i+1)}|^2 + N\sigma^2}, \tag{4.8}$$

and

$$\beta^{(c)} = \frac{(1-K) \sum\limits_{(c-1)R \le i < cR} |e_f^{(i+1)}|^2 / R + N\sigma^2}{2R}. \tag{4.9}$$

As special cases, expressions for the Rayleigh channel and AWGN channel can be obtained by setting $K = 0$ and $K = 1$, respectively. In the Rayleigh channel case, noting that $\alpha^{(c)} = 0$, the general noncentral chi-square distribution of $\mathcal{H}_1$ degenerates to a central chi-square distribution.

In conventional non-compressive spectrum sensing, it is generally valid to assume Gaussianity of noise added on the signal as a prior information. However, in CSS, when Gaussian noise is assumed and certain channel model of original signal, it is noted that

the two hypotheses' distributions in recovered signal do not preserve distributions in the original signal. Specifically, let the sensed spectrum to be $\boldsymbol{x}_f = \boldsymbol{s}_f^r$ in CSS scenario, and it is found that the distributions in (4.7) parameterized by R, (4.8) and (4.9) no longer hold. This finding should not be surprising, as it is a common conclusion that the recovery performance gets worse with decreasing compression ratio and spectrum sparsity, resulting in more occurrences of miss detected and falsely detected spectrum supports. A direct result of this discrepancy is that attempts in CSS to achieve CFAR in detection using threshold setting methods [97] for non-compressive spectrum sensing will lead to varying false alarm probability, which have been appeared in [15, Fig. 7, Fig. 8], [181, Fig. 6, Fig. 8] and [185, Fig. 4]. To give a direct example of such changes on signal statistics, comparisons of original and recovered on spectrum, average channel PSD, and histogram of average channel PSD are given in Fig. 4.2. To simplify the problem, from this point, it is assumed the hypothesis $\mathcal{H}_1$ to have same parameters across all channels, i.e. $\alpha = \alpha^{(1)} = \cdots = \alpha^{(C)}$ and $\beta = \beta^{(1)} = \cdots = \beta^{(C)}$. Hence, the average channel PSDs $p_c^{(c)}$ from different channels are not differentiated. Instead, they are treated as multiple samples of p_c as they are drawn from the same statistical model. The scales of axis in each subfigure pairs for comparison are fixed to give a clear view of the differences. It is particularly noted that the difference in the statistics of p_c between original and recovered signal is obvious in Fig. 4.2 (c) and (f).

4.3 Modelling and Learning of Channel Energy Statistics for Compressive Spectrum Sensing

In this section, the statistical model of channel energy in the CSS is formulated, and an EM-based algorithm to learn the unknown parameters in the formulated model is presented.

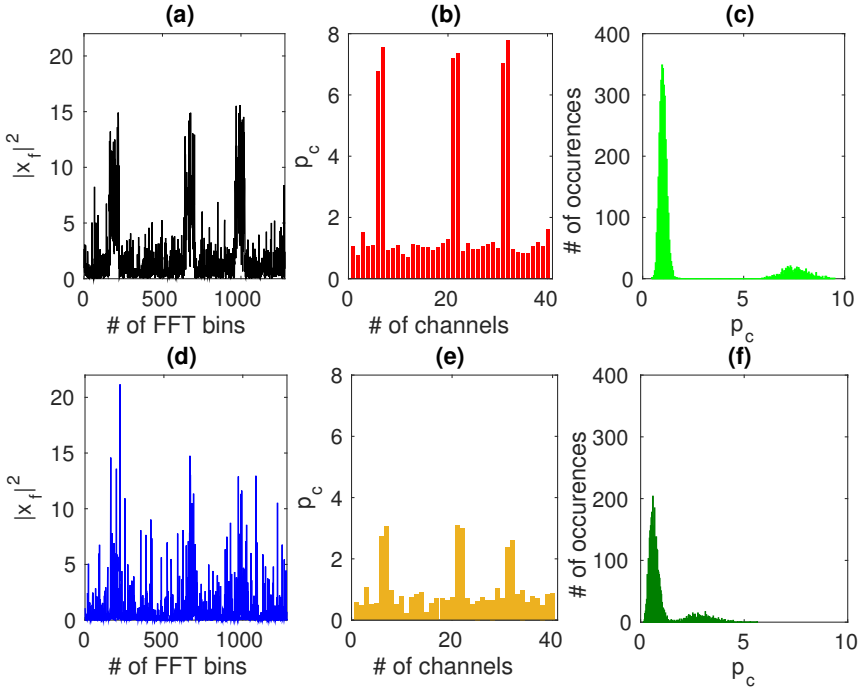


Figure 4.2: Comparisons of original and recovered signal and their statistics. **(a)(d)** PSDs of original and recovered signal, respectively; **(b)(e)** average channel PSDs of original and recovered signal, respectively; **(c)(f)** average channel PSD distributions of original and recovered signal, respectively.

4.3.1 Model and Problem Formulation

To examine the statistics of the recovered signal of compressive spectrum sensing, the concise and direct way is through mathematical analysis of the probability density function of the output signal step by step following certain compressive sensing algorithm. However, common compressive sensing algorithms include optimization sub-routines that have complex and evidently nonlinear forms [55, 91, 94], which makes the PDF derivation rather challenging. Instead, a general statistical model is tentatively postulated where the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ of recovered signal still conform to central and non-central chi-square distribution respectively with parameters relaxed as unknown. To express

this postulated statistical model on the two hypotheses, one has

$$p_c = \begin{cases} \frac{\sigma^2}{2R} r_0,\ r_0 \sim \chi^2(2R) & \mathcal{H}_0 \\ \beta r_1,\ r_1 \sim \chi'^2(2R, \alpha) & \mathcal{H}_1 \end{cases} \tag{4.10}$$

where σ^2, R, α and β are no longer defined as in (4.4), (4.8) and (4.9), and are treated as unknowns which are to be estimated from the learning dataset. This effectively formulates an MM learning problem of two components in each channel. Specifically, after performing multiple runs of CS recovery of the spectrum, suppose there is a series of T recovered spectra as the learning dataset. Thus, one can obtain CT samples of $p_c[t]$ $(t = 1, 2, \cdots, CT)$ according to (4.4). The expression of likelihood function is formulated as

$$\begin{aligned} &f(p_c; \sigma^2, R, \alpha, \beta, \pi_{\mathcal{H}_0}, \pi_{\mathcal{H}_1}) \\ =&\pi_{\mathcal{H}_0} f_{\mathcal{H}_0}(p_c; \sigma^2, R) + \pi_{\mathcal{H}_1} f_{\mathcal{H}_1}(p_c; R, \alpha, \beta), \end{aligned} \tag{4.11}$$

where prior probability is represented as $\pi_{\mathcal{H}_0} = \Pr(Z = \mathcal{H}_0)$ and $\pi_{\mathcal{H}_1} = \Pr(Z = \mathcal{H}_1)$. The MM learning problem mentioned above is to find the optimal $\boldsymbol{\theta}^o$ of parameters

$$\boldsymbol{\theta} = [\sigma^2, R, \alpha, \beta, \pi_{\mathcal{H}_0}, \pi_{\mathcal{H}_1}] \tag{4.12}$$

to obtain maximum likelihood (ML) estimation of parameters, expressed by

$$\begin{aligned} \boldsymbol{\theta}^o &= \arg\max_{\boldsymbol{\theta}} \prod_{t=1}^{CT} f(p_c[t]; \boldsymbol{\theta}) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{CT} \log\left[f(p_c[t]; \boldsymbol{\theta})\right]. \end{aligned} \tag{4.13}$$

Problem (4.13) can be solved by EM algorithm [186] which sets a surrogate majorization function

$$\begin{aligned} Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>}) &= \mathrm{E}_{Z|p_c;\boldsymbol{\theta}^{<j>}}\Big\{\sum_{t=1}^{CT}\log\left[f(p_c[t],z[t];\boldsymbol{\theta})\right]\Big\} \\ &= \sum_{z=\mathcal{H}_0}^{\mathcal{H}_1}\sum_{t=1}^{CT}\Pr(Z=z|p_c=p_c[t];\boldsymbol{\theta}^{<j>})\log\Big\{\pi_z f_z(p_c[t];\boldsymbol{\theta})\Big\} \end{aligned} \tag{4.14}$$

parameterized by $\boldsymbol{\theta}^{<j>}$ and maximize the surrogate function over $\boldsymbol{\theta}$ iteratively. The joint probability density in (4.14) of p_c and latent variable of hypothesis $Z = z$ in (4.14) is expressed as

$$\begin{aligned} f(p_c,z;\boldsymbol{\theta}) =& \pi_{\mathcal{H}_0}\delta(z=\mathcal{H}_0)f_{\mathcal{H}_0}(p_c;\sigma^2,R) \\ &+ \pi_{\mathcal{H}_1}\delta(z=\mathcal{H}_1)f_{\mathcal{H}_1}(p_c;R,\alpha,\beta), \end{aligned} \tag{4.15}$$

where $\delta(a = A)$ is indicator function which equals to 1 only if $a = A$ and elsewhere 0. It is proved [186] that by iteratively decreasing the surrogate function, each EM procedure will not decrease the objective until convergence to at least a local optimum.

In the remaining part of Section 4.3, the iterative maximization of the surrogate function $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>})$ is illustrated, and in the following part of Chapter 4, the superscipt $(\cdot)^{<j>}$ denotes the parameter of the jth iteration.

4.3.2 EM Algorithm: Expectation

In expectation step, the probability term conditioned on $\boldsymbol{\theta}^{<j>}$ in (4.14) is called "membership probability" and obtained by the definition

$$\begin{aligned} M_z^{<j>}[t] :=& \Pr(Z=z|p_c=p_c[t];\boldsymbol{\theta}^{<j>}) \\ =& \frac{\pi_z^{<j>} f_z(p_c[t];\boldsymbol{\theta}^{<j>})}{\pi_{\mathcal{H}_0}^{<j>} f_{\mathcal{H}_0}(p_c[t];\boldsymbol{\theta}^{<j>}) + \pi_{\mathcal{H}_1}^{<j>} f_{\mathcal{H}_1}(p_c[t];\boldsymbol{\theta}^{<j>})}, \end{aligned} \tag{4.16}$$

where $z = \mathcal{H}_0, \mathcal{H}_1$.

4.3.3 EM Algorithm: Maximization

The maximization step of EM procedure finds the next update of parameters $\boldsymbol{\theta}^{<j+1>}$ by setting partial derivatives of $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>})$ to zero if closed-form partial derivatives are available. For prior probabilities, considering the normalization relationship $M_{\mathcal{H}_0}^{<j>} + M_{\mathcal{H}_1}^{<j>} = 1$, the update, which is irrelevant to PDF of chi-square distribution, simply follow

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>})}{\partial \pi_{\mathcal{H}_0}} &= \sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t](\pi_{\mathcal{H}_0})^{-1} - M_{\mathcal{H}_1}^{<j>}[t](1-\pi_{\mathcal{H}_0})^{-1} \\ &= 0 \Rightarrow \pi_{\mathcal{H}_0}^{<j+1>} = \sum_{t=1}^{CT} \frac{M_{\mathcal{H}_0}^{<j>}}{CT}\end{aligned} \tag{4.17}$$

and

$$\pi_{\mathcal{H}_1}^{<j+1>} = 1 - \pi_{\mathcal{H}_0}^{<j+1>}. \tag{4.18}$$

Regarding updating σ^2, it only relates to central chi-square-like PDF $f_{\mathcal{H}_0}$, which has closed-form expression

$$f_{\mathcal{H}_0}(p_c;\sigma^2,R) = \begin{cases} \left(\frac{R}{\sigma^2}\right)^R \cdot \frac{(p_c)^{R-1} e^{-\frac{Rp_c}{\sigma^2}}}{\Gamma(R)}, & p_c > 0 \\ 0, & p_c \leq 0. \end{cases} \tag{4.19}$$

Thus, the partial derivatives of logarithm PDF over σ^2 is not relevant to the Gamma function $\Gamma(\cdot)$, and the update can be presented by,

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>})}{\partial \sigma^2} &= \sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t]\Big(-\frac{R}{\sigma^2} + \frac{p_c[t]R}{\sigma^4}\Big) \\ &= 0 \Rightarrow (\sigma^2)^{<j+1>} = \frac{\sum_{t=1}^{T} M_{\mathcal{H}_0}^{<j>}[t]p_c[t]}{\sum_{t=1}^{T} M_{\mathcal{H}_0}^{<j>}[t]}.\end{aligned} \tag{4.20}$$

Exact optimization of the surrogate function over parameters R, α and β requires differentiating the non-central chi-square PDF $f_{\mathcal{H}_1}$ which is known to have a modified Bessel function term including an infinite series of Gamma function [187], so it is difficult

to derive a straightforward expression of derivatives. The moment-matching method can be adopted as an alternative approach of ML to estimate parameters of MMs by directly matching the moment of mixture's PDF and moment estimations of samples [182]. In this particular case, however, the dimension of parameter vector $\boldsymbol{\theta}$ to be estimated is so large that one needs to match high-order moments and solve high-order equations which is impractical. Although it is unfeasible to directly apply optimization or moment-matching estimation, it is proposed to use simpler moment-matching updates on the above mentioned parameters in the EM's maximization step to get an increase of the surrogate function. Relating to the theory of EM algorithm [186], the non-decreasing property of objective function in (4.13) can be preserved as long as $Q(\boldsymbol{\theta}^{<j+1>};\boldsymbol{\theta}^{<j>}) \geq Q(\boldsymbol{\theta}^{<j>};\boldsymbol{\theta}^{<j>})$, thus the maximization step can be relaxed to an increasing step at a price of possibly slower convergence rate. To illustrate the moment-matching method in the maximization step, to begin with, the update of R is detailed. The second-order central moment of the hypothesis $\mathcal{H}_0$ is expressed as

$$\mathrm{Var}_{\mathcal{H}_0;\boldsymbol{\theta}} = \mathrm{E}\{(p_c - \sigma^2)^2 | Z = \mathcal{H}_0;\boldsymbol{\theta}\} = \sigma^4/R. \tag{4.21}$$

The second-order central moment estimation of samples $p_c[t]$ on condition of $\mathcal{H}_0$ and parameters $\boldsymbol{\theta}^{<j>}$ is expressed as (4.22). Note that a prior occurrence probability of samples $\Pr(p_c = p_c[t])$ is treated as uniform across all t.

$$\begin{aligned}
&\sum_{t=1}^{CT}(p_c[t] - \sigma^2)^2 \cdot \Pr(p_c = p_c[t] | Z = \mathcal{H}_0;\boldsymbol{\theta}^{<j>}) \\
&= \frac{\sum_{t=1}^{CT}\Pr(Z = \mathcal{H}_0 | p_c = p_c[t];\boldsymbol{\theta}^{<j>})[p_c[t] - (\sigma^2)^{<j>}]^2}{\sum_{t=1}^{T}\Pr(Z = \mathcal{H}_0 | p_c = p_c[t];\boldsymbol{\theta}^{<j>})} \\
&= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t][p_c[t] - (\sigma^2)^{<j>}]^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t]}.
\end{aligned} \tag{4.22}$$

Then this second-order central moment estimation (4.22) is used over $\boldsymbol{\theta}^{<j>}$ to match the

parameters given in (4.21) for next update $\boldsymbol{\theta}^{<j+1>}$,

$$\frac{[(\sigma^2)^{<j+1>}]^2}{R^{<j+1>}} = \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t](p_c[t] - (\sigma^2)^{<j>})^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_0}^{<j>}[t]}, \tag{4.23}$$

where $R^{<j+1>}$ is solved with $(\sigma^2)^{<j+1>}$ given in (4.20).

Similarly, given $R^{<j+1>}$, the moment-matching updates for α and β is obtained by the first-order moment and second-order central moment of $\mathcal{H}_1$,

$$\begin{aligned} \mu_{\mathcal{H}_1;\boldsymbol{\theta}^{<j+1>}} &= \beta^{<j+1>}[2R^{<j+1>} + \alpha^{<j+1>}] \\ &= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{<j>}[t]p_c[t]}{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{<j>}[t]}, \end{aligned} \tag{4.24}$$

and

$$\begin{aligned} &\mathrm{Var}_{\mathcal{H}_1;\boldsymbol{\theta}^{<j+1>}} = 4[\beta^{<j+1>}]^2[R^{<j+1>} + \alpha^{<j+1>}] \\ &= \frac{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{<j>}[t]\{p_c[t] - \beta^{<j>}[2R^{<j>} + \alpha^{<j>}]\}^2}{\sum_{t=1}^{CT} M_{\mathcal{H}_1}^{<j>}[t]}. \end{aligned} \tag{4.25}$$

It should be noted that the updates using the moment-matching method do not necessarily decrease $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{<j>})$. In order to guarantee an explicit non-decreasing step, an additional step is proposed where the moment-matching method only updates the corresponding parameter if the value of the surrogate function is not decreased by the moment-matching method. This is accomplished by calculating the surrogate function $Q(\boldsymbol{\theta}^{<j+1>};\boldsymbol{\theta}^{<j>})$ and $Q(\boldsymbol{\theta}^{<j>};\boldsymbol{\theta}^{<j>})$ where the value of central and non-central chi-square PDFs can be well approximated by numerical methods. Moreover, as (4.24) and (4.25) consist of a quadratic equation set, it is noted that there will exist two sets of solutions $[\alpha_1^{<j+1>}, \beta_1^{<j+1>}]$ and $[\alpha_2^{<j+1>}, \beta_2^{<j+1>}]$. For each update, one should choose the set of solutions leading to the greater increase, if there is any, of the surrogate function.

To finalize this section, the proposed EM-based learning algorithm above is summarized as in Algorithm 3.

Algorithm 3 EM-based learning of channel energy statistics in compressive spectrum sensing

Input: CT average channel PSD samples of $p_c[t]$, $t = 1, 2, \cdots, CT$.
Output: Parameters $\boldsymbol{\theta}^{<j>}$ of two hypotheses' PDF in (4.7) optimized by (4.13).

initialize $(\sigma^2)^{<0>} > 0$, $R^{<0>)} > 0$, $\alpha^{<0>} > 0$, $\beta^{<0>)} >= 0$, $1 > \pi_{\mathcal{H}_0}^{<0>} = 1 - \pi_{\mathcal{H}_1}^{<0>} > 0$, $j = 0$.
repeat
  for each hypothesis $z = \mathcal{H}_0$ and $\mathcal{H}_1$ and each sample $t = 0$ to CT **do**
    update membership probability $M_z^{<j+1>}[t]$ as in (4.16)
  end for
  update $(\sigma^2)^{<j+1>}$ as in (4.20)
  update $R^{<j+1>}$ as in (4.23)
  $\boldsymbol{\theta}_{\text{temp0}} \leftarrow \boldsymbol{\theta}^{<j>}$ with element $R^{<j>}$ replaced by $R^{<j+1>}$
  if $Q(\boldsymbol{\theta}_{\text{temp0}}; \boldsymbol{\theta}^{<j>}) < Q(\boldsymbol{\theta}^{<j>}; \boldsymbol{\theta}^{<j>})$ **then**
    $R^{<j+1>} \leftarrow R^{<j>}$
  end if
  update $\pi_{\mathcal{H}_0}^{<j+1>}$ and $\pi_{\mathcal{H}_1}^{<j+1>}$ as in (4.17) and (4.18)
  solve (4.24) (4.25) to get two sets of solutions $[\alpha_1^{<j+1>}, \beta_1^{<j+1>}]$ and $[\alpha_2^{<j+1>}, \beta_2^{<j+1>}]$
  $\boldsymbol{\theta}_{\text{temp1}} \leftarrow \boldsymbol{\theta}^{<j>}$ with elements $\alpha^{<j>}$ and $\beta^{<j>}$ replaced by $\alpha_1^{<j+1>}$ and $\beta_1^{<j+1>}$
  $\boldsymbol{\theta}_{\text{temp2}} \leftarrow \boldsymbol{\theta}^{<j>}$ with elements $\alpha^{<j>}$ and $\beta^{<j>}$ replaced by $\alpha_2^{<j+1>}$ and $\beta_2^{<j+1>}$
  if $Q(\boldsymbol{\theta}_{\text{temp1}}; \boldsymbol{\theta}^{<j>}) > Q(\boldsymbol{\theta}_{\text{temp2}}; \boldsymbol{\theta}^{<j>})$ **then**
    $\alpha^{<j+1>} \leftarrow \alpha_1^{<j+1>}$, $\beta^{<j+1>} \leftarrow \beta_1^{<j+1>}$
  else
    $\alpha^{<j+1>} \leftarrow \alpha_2^{<j+1>}$, $\beta^{<j+1>} \leftarrow \beta_2^{<j+1>}$
  end if
  if $Q(\boldsymbol{\theta}_{\text{temp1}}; \boldsymbol{\theta}^{<j>}) < Q(\boldsymbol{\theta}^{<j>}; \boldsymbol{\theta}^{<j>})$ and $Q(\boldsymbol{\theta}_{\text{temp2}}; \boldsymbol{\theta}^{<j>}) < Q(\boldsymbol{\theta}^{<j>}; \boldsymbol{\theta}^{<j>})$ **then**
    $\alpha^{<j+1>} \leftarrow \alpha^{<j>}$, $\beta^{<j+1>} \leftarrow \beta^{<j>}$
  end if
  $j \leftarrow j + 1$
until $||Q(\boldsymbol{\theta}^{<j>}; \boldsymbol{\theta}^{<j>}) - Q(\boldsymbol{\theta}^{<j-1>}; \boldsymbol{\theta}^{<j>}||_2 <$ a small threshold

4.4 Threshold Adaption via Noise Statistics Estimation in Compressive Spectrum Sensing

Although in Section 4.3 it is presented that the postulated statistical model (4.7), it is noted that Algorithm 3 is impractical in CSS applications due to the following reasons:

1. Algorithm 3 involves multiple calculations of the surrogate function in each iteration, which requires considerable computational effort;

2. As a common drawback of EM algorithms, Algorithm 3 can converge to one of many local maxima. Setting initial values $\boldsymbol{\theta}^{<0>}$ close to the real maxima helps the algorithm converge to the global maximum [188]. In order to find the global maximum, a common practice is that the algorithm should run multiple times with $\boldsymbol{\theta}^{<0>}$ randomly chosen, which adds more computational complexity. Alternatively, human involvement to choose the proper $\boldsymbol{\theta}^{<0>}$ or other initial value selection scheme should be applied;

3. In real-world applications, the sensed channel energy or channel model of each channel is hardly likely to be the same. Hence the samples from active channels would be drawn from differently parameterized hypothesis $\mathcal{H}_1$, which may lead to failure to fit with the postulated statistical model (4.7).

Due to these problems, in this section, a robust and practical threshold adaption scheme via noise statistics estimation is proposed based on the verified statistical model of hypothesis $\mathcal{H}_0$ in (4.7). However, these drawbacks do not affect the usefulness of Algorithm 3. The purpose of Algorithm 3 is not to provide a practical method - it is proposed to verify the postulated statistical model (4.10) and to compute the parameters $\boldsymbol{\theta}$ with various CSS settings.

4.4.1 The Proposed Threshold Adaption Scheme

In this subsection, a threshold adaption scheme which specifically aims to achieve CFAR in the detection phase is presented. According to NP detection theory, CFAR only relates to hypothesis $\mathcal{H}_0$, which means that only parameter learning of $\mathcal{H}_0$ is required. In the following, procedures of the proposed threshold adaption scheme are detailed.

4.4.1.1 Identification of Vacant Channels in the Learning Dataset

Given the learning dataset which comprises T' samples of spectrum recovered by CS algorithm, prior to estimating the parameters relating to $\mathcal{H}_0$, the first step of the proposed threshold adaption is to initially identify these samples of hypothesis $\mathcal{H}_0$. Specifically, given $t = 0, 1, \cdots, T^{'}$ observations of recovered spectrum and consequently $p_c^{(c)}[t]$'s, it is

intended to identify these channels that are free throughout these observations. Due to sparse spectrum usage and rapid observation acquisition, it is a reasonable assumption that the channel occupancy is static with at least one occupied and multiple vacant channels during the acquisition process. This identification problem falls into the category of clustering [189]. The average channel PSD level $p_c^{(c)}$ of these vacant channels conforms to the same distribution; however, other channels have different statistics (usually larger mean and standard deviation) from these of vacant channels. One can exploit this feature and k-means clustering algorithm to identify these vacant channels. In this specific problem, the k-means clustering is conducted in two dimensions - sample means and sample standard deviation - to produce two clusters. K-means is a basic model of clustering which aims to find the clustering solution with minimum intra-cluster distance. In this particular case, it aims to solve the following optimization problem to obtain the vacant channel set $\mathcal{C}$, which writes

$$\mathcal{C} = \arg\min_{\mathcal{X}} \sum_{\mathcal{X}=\mathcal{S},\mathcal{S}\setminus\mathcal{X}} \sum_{k\in\mathcal{X}} \left\| \boldsymbol{d}^{(k)} - \sum_{v\in\mathcal{X}} \boldsymbol{d}^{(v)}/|\mathcal{X}| \right\|_2 , \tag{4.26}$$

where $\mathcal{S} = 1, 2, \cdots, C$ is the set of all channels and the statistics are presented by two-element vector containing sample means and sample standard deviation

$$\boldsymbol{d}^{(c)} = \left(\frac{\sum_{t=1}^{T'} p_c^{(c)}[t]}{T'}, \sqrt{\frac{\sum_{t=1}^{T'} \left[p_c^{(c)}[t] - \frac{\sum_{\tau=1}^{T'} p_c^{(c)}[\tau]}{T'} \right]^2}{T'-1}} \right) . \tag{4.27}$$

Details of the k-means clustering algorithm are not to be given in this thesis and readers are referred to [189, 190].

4.4.1.2 ML Estimation of Parameters in $\mathcal{H}_0$

After these vacant channels are identified, the samples of these channels are used to estimate the parameters R and σ^2 of chi-square-like distribution in (4.19). Denote the number of identified vacant channels as $C' = |\mathcal{C}|$. ML estimator which maximizes the

joint probability of each sample in the dataset for estimation [191] is adopted, expressed by

$$\begin{aligned}(R^o, (\sigma^2)^o) &= \arg\max_{R,\sigma^2} L(\sigma^2, R) \\ &= \arg\max_{R,\sigma^2} \prod_{t=1}^{T'} \prod_{c\in\mathcal{C}} f_{\mathcal{H}_0}(p_c^{(c)}[t]; \sigma^2, R).\end{aligned} \tag{4.28}$$

The maximization[1] is achieved by solving the stationary point of the log likelihood function, where

$$\begin{aligned}\frac{\partial \log\left[L(\sigma^2, R)\right]}{\partial \sigma^2} = \frac{C'T'R}{\sigma^2} - \frac{R\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{\sigma^4} = 0 \\ \Rightarrow (\sigma^2)^o = \frac{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{C'T'}\end{aligned} \tag{4.29}$$

and

$$\begin{aligned}\frac{\partial \log\left[L(\sigma^2, R)\right]}{\partial R} = C'T'[\log R + 1 - \log\sigma^2 - \frac{\Gamma'(R)}{\Gamma(R)}] \\ + \sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} \log p_c^{(c)}[t] - \frac{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}} p_c^{(c)}[t]}{\sigma^2} = 0.\end{aligned} \tag{4.30}$$

$(\sigma^2)^o$ can be easily solved in (4.29). However, R^o is not easy to solve analytically but can be solved numerically. Use Stirling's expansion [192] of the digamma function when R is relatively large

$$\frac{\Gamma'(R)}{\Gamma(R)} = \log(R) - \frac{1}{2R} - \frac{1}{12R^2} + \frac{1}{120R^4} - \frac{1}{252R^6} + \mathcal{O}\left(\frac{1}{R^8}\right), \tag{4.31}$$

and adopt a partial cut-off sum of the series to insert into (4.30). Thus the equation (4.30) can be solved by the Newton-Ralphson method which implements the iteration

$$R^{<j+1>} \leftarrow R^{<j>} - \left.\frac{\partial \log\left[L(\sigma^2, R)\right]/\partial R}{\partial^2 \log\left[L(\sigma^2, R)\right]/\partial R^2}\right|_{R=R^{<j>}}, \tag{4.32}$$

[1]Examine the equations (4.29) and (4.30). It is noted that partial derivatives of $\log\left[L(\sigma^2, R)\right]$ with regards to $R > 0$ and σ^2 both have only one zero point. Hence global optimality is guaranteed.

where the first and second partial derivatives are easy to obtain as it only relates to polynomial of R. The initial value of iteration should be near the root to ensure fast convergence and it can be obtained by moment-matching. Recall (4.21), and the initial value is obtained by

$$R^{<0>} = \frac{C'T'[(\sigma^2)^o]^2}{\sum_{t=1}^{T'}\sum_{c\in\mathcal{C}}[p_c^{(c)}[t] - (\sigma^2)^o]^2}. \tag{4.33}$$

4.4.1.3 Threshold Adaption

NP tests on channel energy are deployed based on the threshold δ

$$p_c^{(c)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \delta, \tag{4.34}$$

where the threshold can be determined by CFAR strategy. Probability of false alarm are found to satisfy the constant complementary cumulative probability density (CCPD) of $\mathcal{H}_0$,

$$P_f = \Pr(p_c^{(c)} > \delta|\mathcal{H}_0) = \frac{\Gamma(R, \frac{R\delta}{\sigma^2})}{\Gamma(R)}, \tag{4.35}$$

where $\Gamma(a, b)$ is the incomplete Gamma function. For given P_f in CFAR, the corresponding threshold δ can be easily found numerically, for example, by binary search method, as CCPD values can be computed[2] and is monotonically decreasing.

4.4.2 Asymptotic Performance of the Proposed Threshold Adaption Scheme

Here the asymptotic performance of the noise energy statistics estimation and threshold adaption is analyzed. The ML estimator (4.28) has the property of an asymptotic minimum-variance and unbiased estimator [191]. Specifically, the distribution of the parameters learned by the proposed ML method asymptotically converges to a Gaussian distribution

$$\left(R^o, (\sigma^2)^o\right) \overset{C'T'}{\sim} \mathcal{N}\left((R, \sigma^2), \boldsymbol{I}^{-1}(R, \sigma^2)\right), \tag{4.36}$$

[2]Note R in (4.35), as an output of the proposed ML estimation, can be arbitrary positive value. In simulations the Matlab functions `chi2cdf` is used to calculate CCPD.

where Fisher information matrix $\boldsymbol{I}$ is

$$\begin{aligned}\boldsymbol{I}(R,\sigma^2) &= \begin{bmatrix} -\mathrm{E}\left[\frac{\partial^2 \log\left[L(\sigma^2,R)\right]}{\partial R^2}\right] & -\mathrm{E}\left[\frac{\partial^2 \log\left[L(\sigma^2,R)\right]}{\partial R \partial \sigma^2}\right] \\ -\mathrm{E}\left[\frac{\partial^2 \log\left[L(\sigma^2,R)\right]}{\partial \sigma^2 \partial R}\right] & -\mathrm{E}\left[\frac{\partial^2 \log\left[L(\sigma^2,R)\right]}{\partial (\sigma^2)^2}\right] \end{bmatrix} \\ &= \begin{bmatrix} C'T'\left(\frac{1}{R} - \frac{\Gamma''(R)}{\Gamma'(R)}\right) & 0 \\ 0 & \frac{C'T'R}{\sigma^4} \end{bmatrix}.\end{aligned} \tag{4.37}$$

The variance of estimated parameters $\left(R^o, (\sigma^2)^o\right)$ asymptotically approaches Cramer-Rao lower bound,

$$\mathrm{Var}(R^o) \overset{C'T'}{\sim} \frac{1}{C'T'}\left(\frac{1}{R} - \frac{\Gamma''(R)}{\Gamma'(R)}\right)^{-1}, \tag{4.38}$$

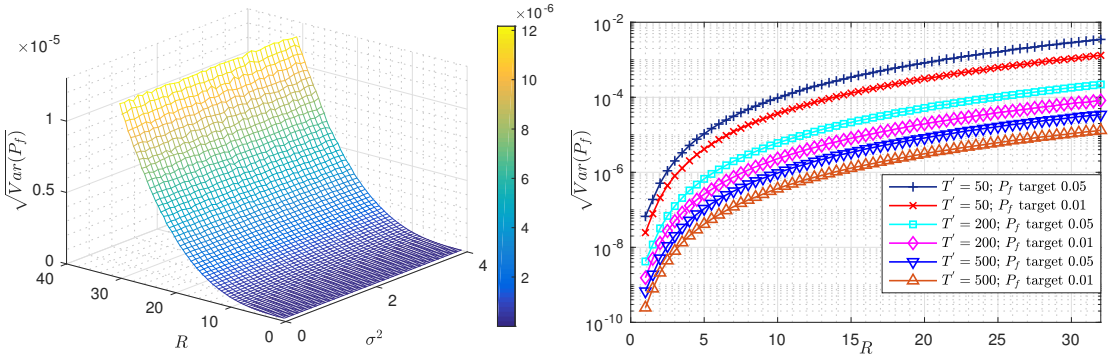
and

$$\mathrm{Var}\left((\sigma^2)^o\right) \overset{C'T'}{\sim} \frac{\sigma^4}{C'T'R}. \tag{4.39}$$

With the increasing and sufficiently large number of samples $C'T'$, these lower bounds (4.38) and (4.39) can be approached asymptotically and well approximates the real variance, which represents the analytic performance of the ML estimation (4.28).

Moreover, the variance of the real P_f is another relevant parameter, as it indicates how well the proposed noise statistics estimation and threshold adaption can determine a threshold to keep a target P_f. To quantify the variance of P_f, the asymptotic variance bounds (4.38) and (4.39) are used. To obtain the closed-form expression of the variance of P_f, the variance of P_f with regards to random variables R^o and $(\sigma^2)^o$ should be derived using the relationship (4.35). As the expression (4.35) has a complex formulation which contains the incomplete Gamma function, the derivation of the variance of P_f is not a trivial task. Alternatively, Monte-Carlo simulations are used to inspect this variance, where the number of null channels is set as $T' = 34$ (the same value are used in Section 4.5 where 6 channels out of 40 are active) and the sample variance in Monte-Carlo simulation is used as the approximation to the real variance. The legitimacy behind this Monte-Carlo simulation is the law of large numbers - for the sufficient simulations, the

sample variance of multiple independent observations of P_f converges to the variance of the random variable P_f. The process of such a Monte-Carlo simulation is detailed as follows. Firstly, choose a target P_f and the true parameters R and σ^2. For each run of the Monte-Carlo simulation, use formula (4.35) with random $R^o \sim \mathcal{N}(R, \mathrm{Var}(R^o))$ and $(\sigma^2)^o \sim \mathcal{N}(\sigma^2, \mathrm{Var}((\sigma^2)^o))$ whose variations are as shown in (4.38) and (4.39), to get a threshold value δ. Next, for each time of simulation, use this obtained δ and the true values of R and σ^2 to get the practical value of P_f. Finally, calculate the (sample) variance of the P_f's resulted in each time of the Monte-Carlo run. Fig. 4.3 (a) gives the numerical results of the standard deviation, as the squared root of the variance of P_f, $\sqrt{\mathrm{Var}(P_f)}$, against R and σ^2 for target $P_f = 0.01$ given $T^{'}$ value of 500. At first glance, it is noted that the variance of P_f is nearly constant across the parameter σ^2's range and even for larger ranges of σ^2 which are not shown. This should not come as a surprise, as one can note that σ^2 from in (4.10) is a factor only accounting for the dilation of the channel energy, which does not affect any statistics of the null hypothesis $\mathcal{H}_0$ expect for a constant scaling value relating to σ^2. Therefore, the value of σ^2 should not affect the asymptotic performance of the proposed threshold adaption scheme. Without the need to consider the varying value of σ^2, in Fig. 4.3, the standard deviation of P_f versus the degree-of-freedom-related parameter R is examined, with fixed value of $\sigma = 1$, for different target P_f values of 0.01 and 0.05 and $T^{'}$ values of 50, 200 and 500. With the interested range of $R \in (0, 32]$ in Fig. 4.3 (b), for fixed sample size $T^{'}$ and a certain target P_f, the variance of practical P_f increases with larger R value, i.e. larger DoF of the chi-squared model of $\mathcal{H}_0$. However, it is highly undesirable that the standard deviation of the practical P_f becomes close to the corresponding target P_f value, which implies a poor ability to keep the practical P_f near the target. By selecting a larger size of the observed dataset $T^{'}$, smaller variations of practical P_f can be achieved. For example, as in Fig. 4.3 (b), the standard deviation of practical P_f can be decreased by more than an order of magnitude, by increasing $T^{'}$ 4 times from 50 to 200, and around two orders of magnitude by increasing $T^{'}$ 10 times from 50 to 500.



(a) (b)

Figure 4.3: Asymptotic lower bound of standard deviation of practical P_f for the proposed threshold adaption scheme. **(a)** the standard deviation of practical P_f versus the two parameters, R and σ^2, of the ML estimation, for sample size $T^{'} = 500$, and target $P_f = 0.01$. **(b)** the standard deviation of practical P_f versus R for various sample size $T^{'}$'s and target P_f's, where $\sigma^2 = 1$ is set.

4.5 Numerical Simulation

In this section, simulation results of the proposed signal energy statistics algorithm and threshold adaption via noise statistics estimation are presented. Time-domain signals are generated as in (4.40) to emulate signals with continuous support in the frequency domain,

$$s_t{}^{(n)} = \sum_{i=1}^{\mathcal{K}} \sqrt{E_i B}\text{sinc}[B(n - n_i)]e^{j2\pi f_i t}, \tag{4.40}$$

where $\mathcal{K}$ is the number of active channels; B is channel bandwidth; E_i, n_i, f_i stands for total energy, time offset and central frequency of the ith active channel. To simulate the band of TVWS in UK (470-790MHz), the interested spectrum bandwidth is chosen to be 320MHz and set $N = 1280$, $B = 8$MHz, $\mathcal{K} = 6$, and $E_i = 1280/6$. For each combination of recovery algorithm, SNR and compression ratio, learning dataset is generated by the channel occupancy patterns as $[f_1, \cdots, f_6] = [36, 44, 164, 172, 244, 252]$(MHz). SNR is defined as SNR $= ||\boldsymbol{s}_t||_2^2/(N\sigma^2)$ where σ^2 is the power of complex zero-mean additive noise $w_t^{(n)} \sim \mathcal{CN}(0, \sigma^2)$.

4.5.1 Effectiveness of Channel Energy Statistics Learning for Compressive Spectrum Sensing

In this subsection, the proposed channel energy statistics learning algorithm is focused on and implemented using the samples $p_c[t]$ from channel set whose central frequencies are $[f_1, \cdots, f_6]$ and $[f_7, \cdots, f_{12}] = [60, 68, 188, 196, 268, 276]$(MHz) so that this set of $C = 12$ channels has channel occupancy rate of 0.5. The dimension of the dataset is $T = 300$.

The proposed learning algorithm is experimented over recovered signals by three major sparse recovery algorithms - BP, OMP and SBL, and two propagation scenarios - the AWGN and Rayleigh channel. Fig. 4.4 (a) and (c) exemplify the distributions of average channel PSD level $p_c^{(c)}$ conditioned on hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ and unconditional distribution whose parameters are learned from the proposed algorithm, using BP, a compression ratio of 0.2 and SNR of 0dB, over the AWGN and Rayleigh channels respectively. It can be seen that the learned distributions align well with the histogram of the learning dataset in both cases. Fig. 4.4 (b) and (d) correspond to (a) and (c) respectively and show the increasing and convergent results of likelihood functions which the proposed learning algorithm aims to maximize. Fig. 4.5 (a) and (b) show the learning results of signals recovered by OMP and SBL algorithm respectively. Again, SNR is set to 0dB in AWGN channel and compression ratio of 0.5 and 0.25 in (a) and (b) respectively. The examples in Fig. 4.5 indicate that the learning results may also describe the empirical distribution of learning dataset with OMP and SBL recovery algorithms.

To further characterize how well the empirical distribution from simulation matches the postulated model in (4.10), the Kullback-Leibler divergence (KLD) is adopted as a measure of the similarity of the two distributions [193]. The KLD may be any non-negative value KLD of 0 reveals two identical distributions, and a smaller KLD value towards 0 indicates more similarity. From simulations, the learning dataset forms a histogram of empirical distribution, instead of a continuous probability density function.

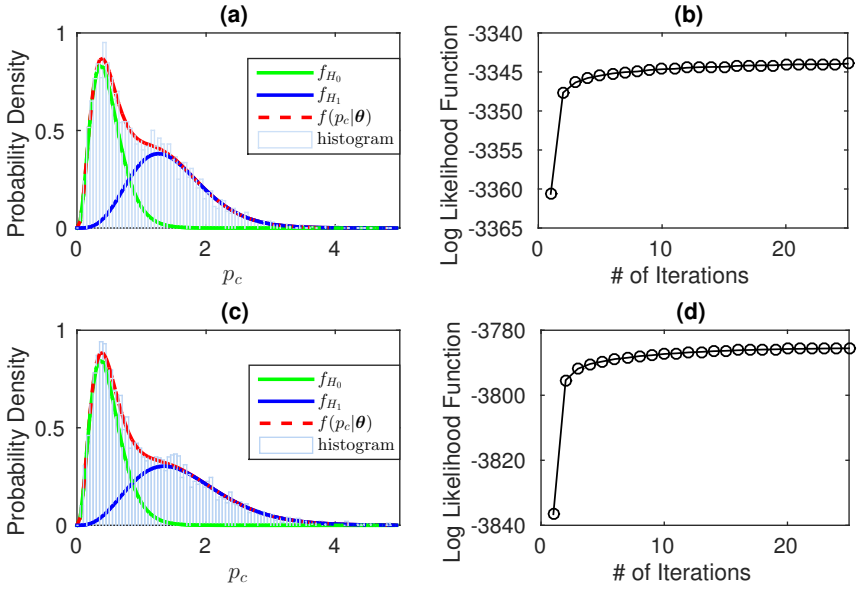


Figure 4.4: **(a)(c)** Histogram and learned distributions of BP recovery, SNR = 0dB, compression ratio of 0.2, in AWGN and Rayleigh channel respectively; **(b)(d)** Likelihood function over first 25 iterations of learning process in (a) and (c), respectively.

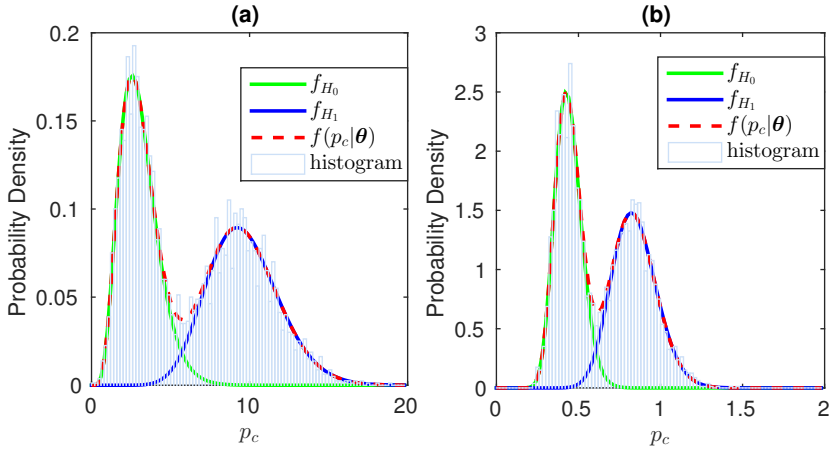


Figure 4.5: Histogram and learned distributions, SNR = 0dB in AWGN channel, using **(a)** OMP with compression ratio of 0.5 and **(b)** SBL with compression ratio of 0.25.

Therefore, the discrete probability form of the KLD definition is adopted,

$$D_{\mathrm{KL}}(P_{\mathrm{empirical}} \| P_{\mathrm{model}}) = \sum_{i=1}^{I} P_{\mathrm{empirical}}(i) \log \left(\frac{P_{\mathrm{empirical}}(i)}{P_{\mathrm{model}}(i)} \right), \tag{4.41}$$

where $P_{\text{empirical}}(i)$ is the histogram value of empirical distribution, normalized by

$$\sum_{i=1}^{I} P_{\text{empirical}}(i) = 1; \tag{4.42}$$

$P_{\text{model}}(i)$ is the normalized probability in the ith interval of the postulated model, expressed by

$$P_{\text{model}}(i) = \frac{\int_{i\text{th interval}} f_z(p_c;\boldsymbol{\theta}^o)\mathrm{d}p_c}{\int_{\text{all intervals in } \mathcal{I}} f_z(p_c;\boldsymbol{\theta}^o)\mathrm{d}p_c},\ z = \mathcal{H}_0, \mathcal{H}_1. \tag{4.43}$$

The range of interested intervals is determined as

$$\begin{aligned}[a,b] = \Big[&\min\left\{p_c\left[t\right]|t = 1,2,\cdots,T^{'}\right\}, \\ &\max\left\{p_c\left[t\right]|t = 1,2,\cdots,T^{'}\right\}\Big].\end{aligned} \tag{4.44}$$

Then the interval is evenly divided into I (in simulations $I = 250$ is used) smaller ones such that the ith interval corresponds to $\left[a + (i-1)\frac{b-a}{I}, a + i\frac{b-a}{I}\right]$. Fig. 4.6 plots the KLD values from empirical data to the postulated model, for the three algorithms, against compression ratio and SNR respectively. For a small KLD close enough to 0, it can be understood that the two distributions are nearly identical with minor similarities. In all cases of the simulations, KLD values turn out to be small, within the range from 10^{-2} to 5×10^{-2}. With these small KLD values, the postulated models with the learned parameters are verified to be good representations of the corresponding empirical distributions of the channel energy values. It is also noted that in Fig. 4.6 the KLD values tend to be relatively higher for smaller compression ratios and SNRs, and also for the OMP algorithm compared to the other two algorithms. This increased KLD values can be preliminarily explained by the fact that under these circumstances the recovery algorithms tend to have poorer recovery performance and the resultant mixture models of channel energy statistics are less evident, i.e. the distributions of two hypotheses are less separated, so that the EM parameter learning procedure has more propensity to find a local minimum that deviates from the true model.

Next, the compression ratio is made varying and the parameter changes learned

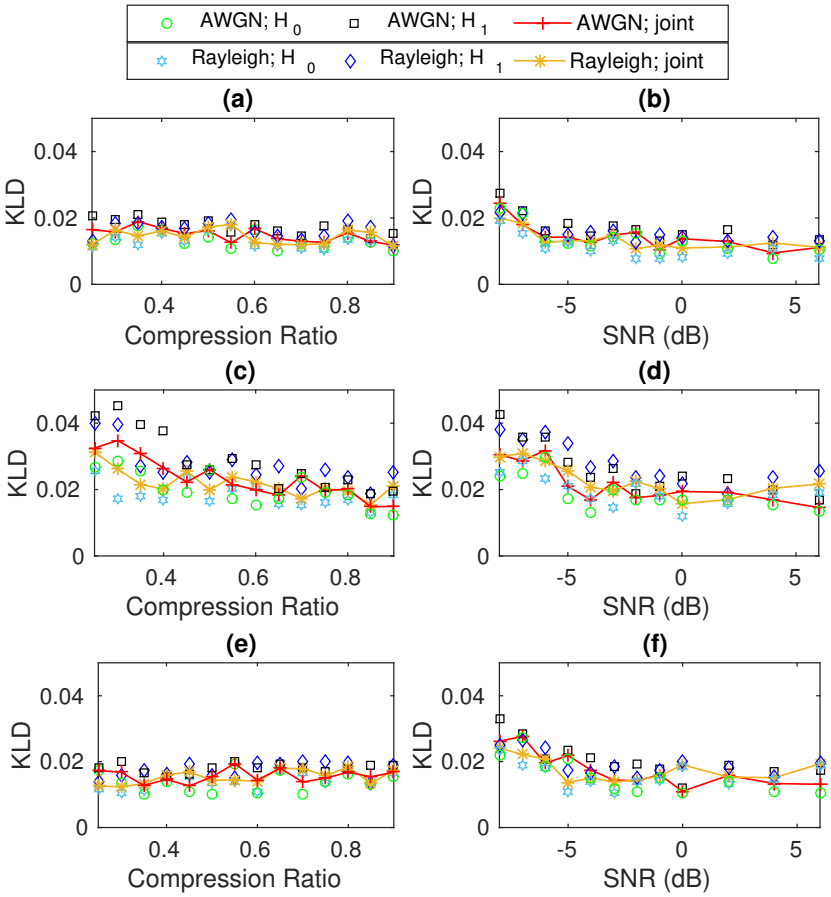


Figure 4.6: KLD from empirical data to postulated models versus compression ratio and SNR. **(a)(b)** BP; **(c)(d)** OMP; **(e)(f)** SBL.

by the proposed algorithm are focused on. Fig. 4.7 gives these results of three recovery algorithms in AWGN channel under the condition SNR = 0dB. Moreover, each parameter value of the average channel PSD level distribution of the original signal is drawn as a reference. As a general and intuitive trend, most parameters from signals recovered by three algorithms approach to the original signal's reference with increasing compression ratio. An interesting exception is R of SBL which fluctuates slightly near 32. This can be explained by SBL's assumption of Gaussian prior distribution of the signal to be recovered [94]. Given this assumption and by the definition of the chi-square distribution, the distribution of average channel PSD level has a definite DoF which equals twice the number of PSD bins in each channel, considering complex Gaussian noise over each PSD bin. Fig. 4.8 gives similar results in Rayleigh channel situation with the same settings

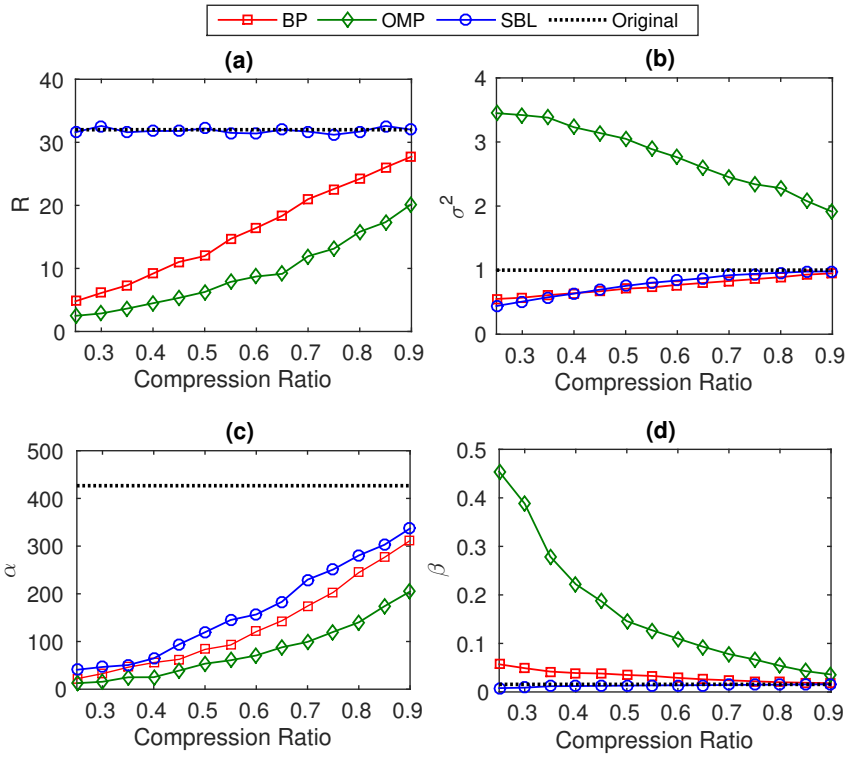


Figure 4.7: Learned parameters against compression ratio using three major recovery algorithms, compared with original signal parameters. SNR = 0dB, AWGN channel.

as in Fig. 4.7. It is noted that the SBL's parameter α is negligible and approximates the original signal's reference of 0. In the Rayleigh channel scenario, not only the elements in $\boldsymbol{s}_f$ corresponding to the hypothesis $\mathcal{H}_0$ but also $\mathcal{H}_1$ have Gaussian distribution. This prior distribution of original signals matches the SBL's Gaussian assumption. As a result, SBL induces the Gaussianity of the original signal set, which leads to central chi-square statistics in $\mathcal{H}_1$ in the recovered signal.

4.5.2 Performance of Noise Energy Estimation and Threshold Adaption

In this section, the performance of the noise energy estimation and threshold adaption proposed in Section 4.4 are evaluated. As the first step, k-means clustering is performed for several learning datasets with different SNRs and dataset dimensions. Qualitatively, it is obvious from CLT that the larger the dataset dimension is, the more concentrated

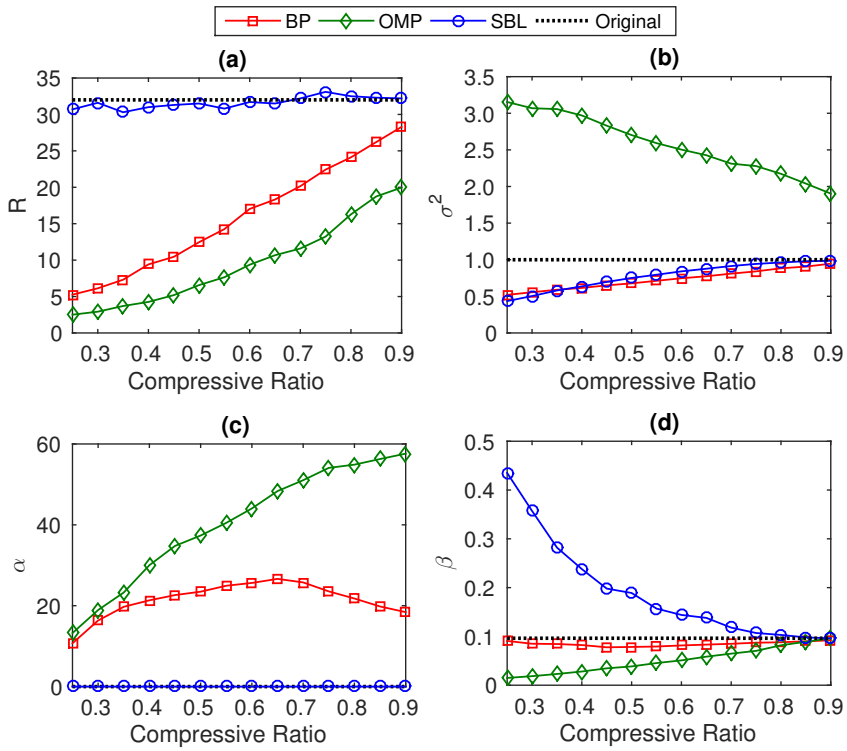


Figure 4.8: Learned parameters against compression ratio using three major recovery algorithms, compared with original signal parameters. SNR = 0dB, Rayleigh channel.

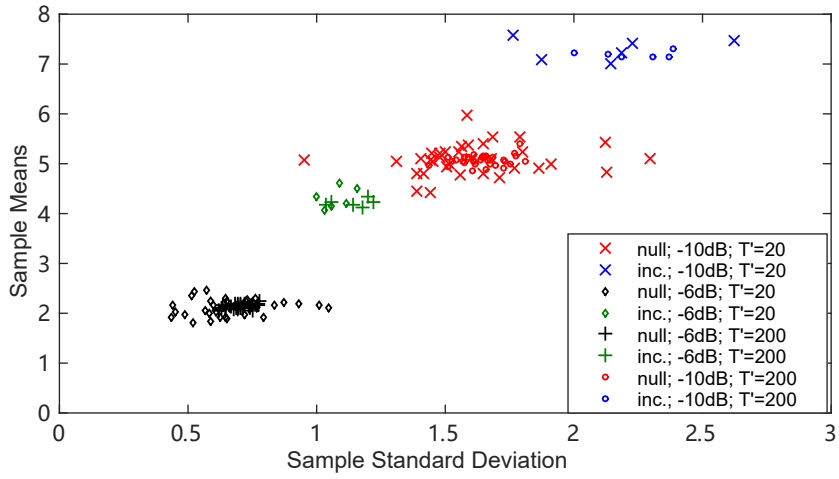


Figure 4.9: K-means clustering for identifying the vacant channels ('null' in the legend) from incumbent ones ('inc.' in the legend). It is shown that the clustering algorithm successfully identifies the vacant channels with low SNR (-10dB) and small dataset ($T' = 20$).

the statistics $\boldsymbol{d}^{(c)}$ within one cluster (for either vacant or incumbent channels) tend to be. Thus, it means that the global maximum of clustering objective (4.26) is more likely to correspond to the true channel occupancy pattern. On the other hand, in lower SNR cases, the statistics $\boldsymbol{d}^{(c)}$ from vacant and incumbent channel clusters are less separated. To investigate the effectiveness of the proposed k-means clustering step, the clustering results are presented with a small dataset ($T^{'} = 20$) and low SNR (-10dB), accompanied by other three clustering results with larger dataset $T^{'} = 200$ and higher SNR of -8dB. Although the clustering performance tends to deteriorate for decreasing the SNR and the sample size $T^{'}$, the results in Fig. 4.9 show that the proposed k-means clustering suffices to identify the vacant channels in a worst case with $SNR = -10$dB and small dataset size $T^{'} = 20$. Here, the SNR of -10dB is a level where compressive spectrum recovery becomes unreliable so that the detection performance is not usable.

Next, the samples of $p_c^{(c)}$ from identified vacant channels are extracted, and ML estimation based on $T^{'} = 200$ spectrum samples are performed and the thresholds δ given $P_f = 0.01$ are obtained. In Fig. 4.10 and Fig. 4.11, the adapted thresholds are plotted for three major recovery algorithms, against compression ratio and SNR respectively. To stress the necessity of the proposed threshold adaption scheme, again, thresholds for non-compressive cases are calculated based on original signals $\boldsymbol{s}_t$ and drawn for reference. These two figures are direct illustrations of how evidently discrepant the thresholds should be valued among varying SNRs, compression ratios, and different choices of recovery algorithms, and especially between non-compressive and compressive cases. Moreover, as an intuitive rule, thresholds of all three CS algorithms approach those of non-compressive cases with compression ratio and SNR increasing.

Using the thresholds adapted, Monte-Carlo tests of the detection performance are performed by using these thresholds for detection over a new sample set of the recovered spectrum. Note that the dataset used here for detection performance simulation is other than the learning dataset used for threshold adaption. Fig. 4.12 illustrates the probability of detection and false alarm against compression ratio using the proposed threshold

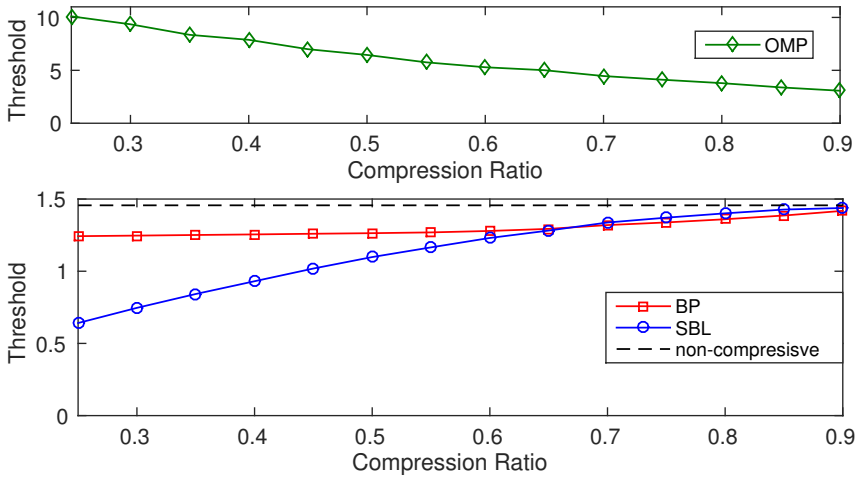


Figure 4.10: Threshold adapted from the proposed scheme against compression ratio of three major recovery algorithms, given target $P_f = 0.01$. SNR = 0dB.

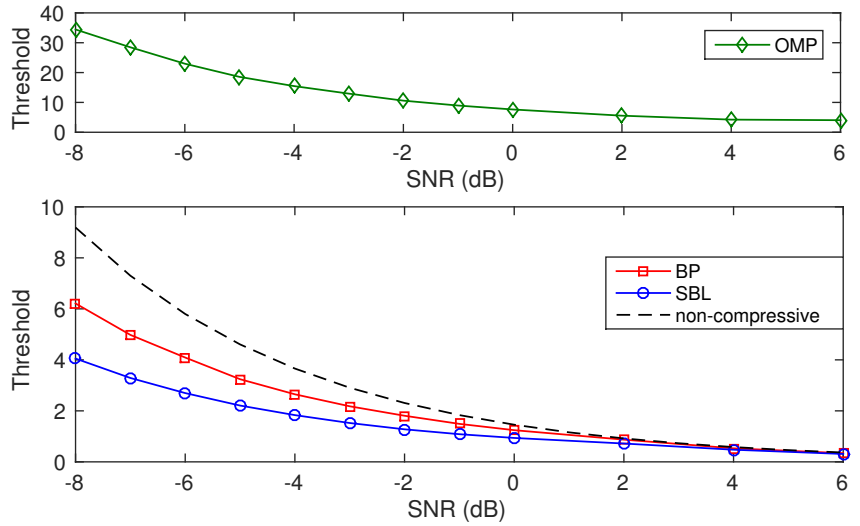


Figure 4.11: Threshold adapted from the proposed scheme against SNR of three major recovery algorithms, given target $P_f = 0.01$. Compression ratio is 0.4.

adaption scheme for the three major recovery algorithms with SNR = 0dB. It is discovered that the proposed threshold adaption scheme can maintain a close approximate of predefined P_f of 0.01. Similarly, Fig. 4.13 gives CFAR detection performance against SNR for the three selected recovery algorithms while keeping the compression ratio 0.4. Again, the actual P_f tends to maintain constant with minor variance. Moreover, the results of P_f which are kept nearly constant from Fig. 4.12 and Fig. 4.13 in turn verify

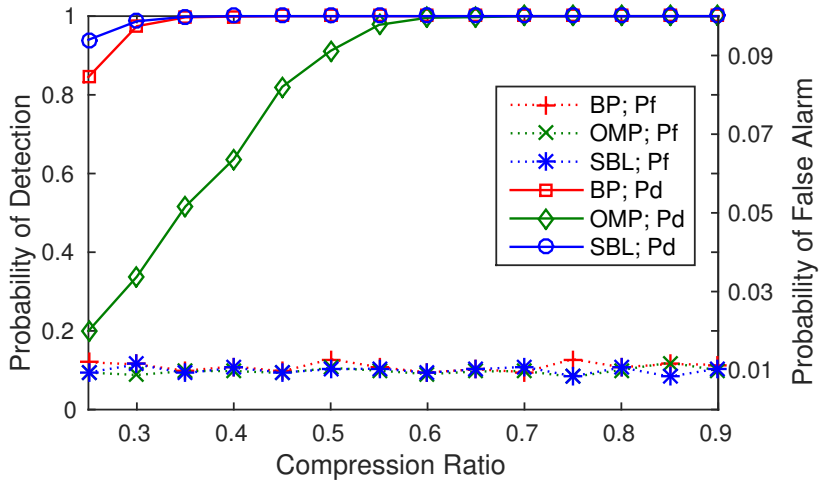


Figure 4.12: CFR detection performance of three major recovery algorithms against compression ratio using proposed noise energy statistics estimation and threshold adaption. SNR = 0dB.

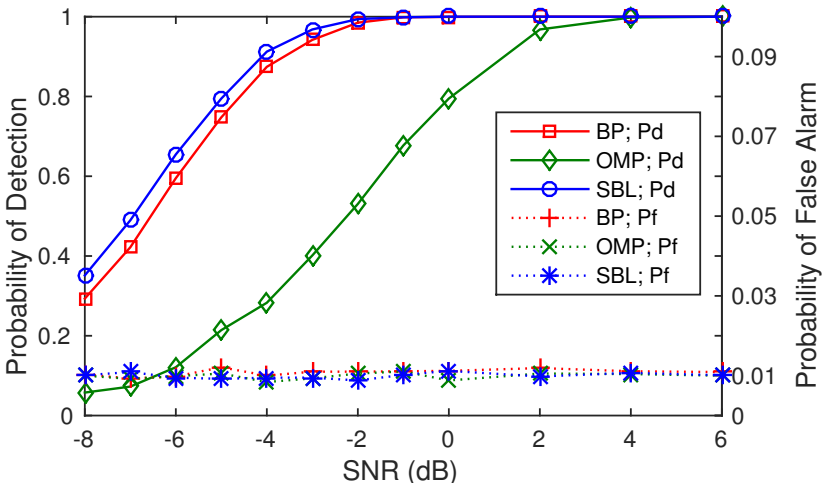


Figure 4.13: CFR detection performance of three major recovery algorithms against SNR using proposed noise energy statistics estimation and threshold adaption. Compression ratio is 0.4.

the reasonableness of the postulated model (4.10) and the effectiveness of the proposed threshold adaption scheme via noise energy statistics estimation.

4.6 Summary

In this chapter, the inconsistency of channel energy statistics and hence optimal threshold between CSS and conventional non-compressive spectrum sensing is addressed for the first time. Then a channel energy statistics model was postulated for CSS and an algorithm used for learning the parameters of the postulated model was proposed. It is postulated that the two hypotheses of channel energy of absent and active radio activity conform to central and noncentral chi-square distributions, respectively, both with unknown parameters. Given the proposed statistics model being validated, a robust and practical scheme of threshold adaption achieved by channel noise statistics estimation was proposed. The asymptotic performance of the scheme was analyzed by presenting the variance lower bound of actual false alarm probability. By numerical simulations, it was shown that the postulated model provided a good fit to the empirical distributions, in the sense of data-to-model KLD. Next, out of interest for the trend of these parameters in different CS settings, the learning algorithm is experimented with in various compression ratio settings and presented the trend of each parameter versus compression ratio. Furthermore, the adapted thresholds given target false alarm probability were plotted against SNR and compression ratio, where the discrepancies in thresholds between compressive and non-compressive spectrum sensing are directly illustrated. Moreover, the actual false alarm probability using the thresholds adapt by the proposed scheme was shown to be near the target value, which inversely verified the validity of the postulated model.

Chapter 5

Compressive Spectrum Sensing based on Practical Multicoset Sampler towards mmWave

Compressive sensing (CS) has been introduced to be applied to wideband spectrum sensing to achieve real-time sampling at sub-Nyquist rates, hence alleviating the need for the complexity and high cost of analog-to-digital converters (ADCs) [15, 161]. Several practical baseband processing architectures have been proposed in the literature to obtain the compressed signal under the frame of CS, including the random demodulator [71], the multicoset sampler [61] and the modulated wideband converter (MWC) [65]. The multicoset sampler has a simpler hardware architecture than the MWC, and effectively constructs a multiple-measurement-vector (MMV) model of CS, instead of single-measurement-vector (SMV) produced by the random demodulator. Inheriting the configurations of signal reconstruction system based on interleaved ADCs [194], it employs multiple parallel ADCs sampling various time-shifted versions of the baseband signal with a fraction of the Nyquist rate. Although the concept of the ideal multicoset sampler has been widely discussed, there is little in the literature on the real-world mod-

eling and implementation of the multicoset sampler. It is hence important to identify the challenges and potentials to implement the multicoset-sampler-based compressive spectrum sensing for wideband scenarios, especially for multi-gigahertz baseband for promising millimeter-wave (mmWave) applications.

Practical attempts have been known to realize a 4-branch MWC front-end and the CSS system for the instantaneous bandwidth of 6 GHz, working with National Instruments (NI) software-defined radio (SDR) platform [81]. With the four ready-to-use analog-to-digital converters (ADCs) sampling at 120 MHz equipped on the SDR platform, the major design challenge lies in the front end subsystem, where the generation, distribution, and synchronization of the 6 GHz-chip-rate pseudorandom sequences, which has been shown to be implemented using a field-programmable gate array (FPGA) development board, ensuring tight synchronization and low phase noise [82]. Moreover, such an MWC system's implementation is designed to processing real-valued signals, not suitable for in-phase-and-quadrature processing for general baseband applications.

One major reason why the multicoset sampler architecture is of particular interest in the quest of real-world realization as its comparatively straightforward architecture makes it more feasible to be built using off-the-shelf ADCs without introducing much hardware complexity that might undermine the advantage of sub-Nyquist sampling. The implementation of the random demodulator has been largely discussed by the integrated-circuit design communities, as the wideband analog integrator before the sampling-and-hold device is the key component that has to be built as application-specific. Looking at the MWC, although it is possible to be implemented using general-purpose devices, in addition to the ADC array, one can note that the parallel structure of MWC requires more stages of analog processing, e.g. a low pass filter and a mixer, inducing more hardware complexity and uncontrollable factors on the consistency of the analog chains and possibly impeding the expansion to more analog branches. Whereas for the multicoset sampler, the main design issue is how to realize the random and precise timing offsets in analog branches or in the timing of the sampling, which already has a good technical

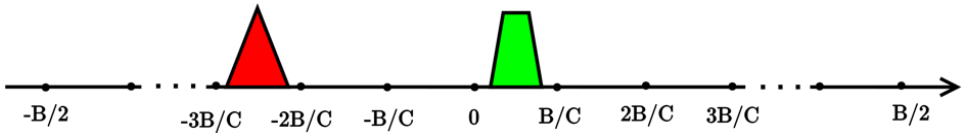


Figure 5.1: Pictorial illustration of the interested spectrum and channels.

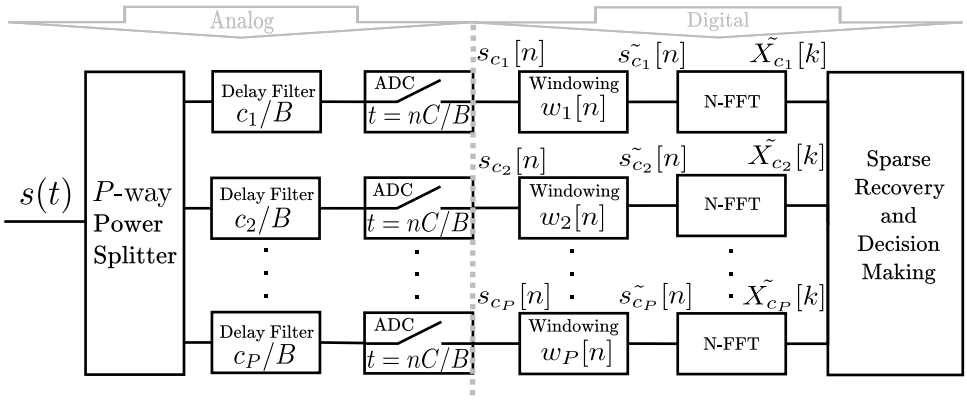


Figure 5.2: The proposed baseband processing model of multicoset sampler.

foundation in a similar setting of the interleaved ADCs.

Given the aforementioned potentials of the multicoset sampler architecture, this chapter is motivated by its real-world implementations, starting with the modeling of the practical multicoset sampler in Section 5.1 and then the important issue of spectrum leakage in Section 5.2. Sparsity estimation and dimension reduction are proposed to be applied to the practical multicoset model to enable greedy algorithms and to further reduce the recovery complexity in Section 5.3. Then, a CSS demo at 28.5 GHz with 3.072 GHz bandwidth is presented based on a mmWave SDR platform to illustrate the complete design of a multicoset-sampler-based CSS. Finally, it is identified that the precision of these offsets, which is limited in practice due to imperfection, has impacts on the precision of the compressed measurements as well as the recovered spectrum. Such measurement errors induced by undesired timing skews are analyzed and quantified in Section 5.5.

5.1 System Model of Multicoset Sampler

Denote a bandlimited complex time-domain signal by $s(t) \in \mathbb{C}, t \in \mathbb{R}$ spanning the frequency band of $\left[-\frac{B}{2}\ (\text{Hz}), \frac{B}{2}\ (\text{Hz})\right]$, where there are C equally-spaced channels with the bandwidth of B/C (Hz), as illustrated in Fig. 5.1. A multicoset sampler [16, 65] with P ($P < C$) cosets discretize the analog signal $s(t)$ at a fraction-C of the Nyquist sampling rate $\frac{1}{B}$ (Hz), which writes

$$s_{c_p}[n] = s\left(\frac{nC - c_p}{B}\right), \quad n \in \mathbb{Z} \tag{5.1}$$

for $p = 1, 2, \cdots, P$, where $c_p \in [0, C)$ denotes the ratio of time delay of the pth coset to the Nyquist-rate period. The frequency-domain presentation, i.e. discrete-time Fourier transform (DTFT) of the sampled signal $s_{c_p}[n]$'s is to be examined by definition $X_{c_p}\left(e^{j2\pi f\frac{C}{B}}\right) := \sum_{n=-\infty}^{+\infty} s_{c_p}[n] \exp\left(-j2\pi f nC/B\right)$. It is a convenient treatment that the DTFT of each coset's signal is to be expressed by a Fourier transform of a series of Dirac delta pulses

$$\begin{aligned} X_{c_p}\left(e^{j2\pi f\frac{C}{B}}\right) &:= \sum_{n=-\infty}^{+\infty} s_{c_p}[n] \exp\left(-j2\pi f n\frac{C}{B}\right) \\ &= \int_{-\infty}^{+\infty} s(t - \frac{c_p}{B}) \cdot \left[\sum_{n=-\infty}^{+\infty} d(t - \frac{nC}{B})\right] \exp\left(-j2\pi f t\right) \mathrm{d}t, \end{aligned} \tag{5.2}$$

where $d(\cdot)$ denotes Dirac delta function and $\sum_{n=-\infty}^{+\infty} d(t - \frac{nC}{B})$ is a Dirac comb whose Fourier transform is

$$\mathcal{F}\left\{\sum_{n=-\infty}^{+\infty} d(t - \frac{nC}{B})\right\} = \frac{B}{C}\sum_{n=-\infty}^{+\infty} d(f - \frac{nB}{C}). \tag{5.3}$$

By rewriting (5.2) and applying (5.3), the DTFT of each coset's signal can be related to the Fourier transform of $s(t)$ denoted by $X(f) := \mathcal{F}\left\{s(t)\right\} = \int_{-\infty}^{+\infty} s(t) \exp(-j2\pi f t)\mathrm{d}t$.

It is expressed by

$$\begin{aligned} X_{c_p}\left(e^{j2\pi f\frac{C}{B}}\right) &= \mathcal{F}\left\{s(t-\frac{c_p}{B})\cdot\sum_{n=-\infty}^{+\infty}d(t-\frac{nC}{B})\right\} \\ &= \left[X(f)\exp\left(-j2\pi c_p\frac{f}{B}\right)\right] \underset{\text{(convolution)}}{*} \left[\frac{B}{C}\sum_{n=-\infty}^{+\infty}d(f-\frac{nB}{C})\right] \\ &= \frac{B}{C}\sum_{n=-\infty}^{+\infty}\exp\left(j2\pi c_p\left(\frac{n}{C}-\frac{f}{B}\right)\right)X\left(f-\frac{nB}{C}\right) \\ &= \frac{B}{C}\exp\left(-j2\pi c_p\frac{f}{B}\right)\sum_{n=-\infty}^{+\infty}\exp\left(j2\pi c_p\frac{n}{C}\right)X\left(f-\frac{nB}{C}\right) \end{aligned} \tag{5.4}$$

Note that the DTFT in (5.4) by definition is $\frac{B}{C}$-periodic, and one only needs to examine one period with regards to f. To align the channel structure of the shifted spectrum in (5.4) with the interested interval, one can choose any $l \in \mathbb{Z}$ to make the interval $f \in \left[\frac{lB}{C}, \frac{(l+1)B}{C}\right)$. Recall $X(f) = 0$ for $f \notin \left[-\frac{B}{2}, \frac{B}{2}\right]$, the number of summations in (5.4) is reduced to C, and (5.4) is rewritten for $f \in \left[0, \frac{B}{C}\right)$ as

$$\begin{aligned} Y_{c_p}(f) &= \sum_{n=-\frac{C}{2}+1}^{\frac{C}{2}}\exp\left(j2\pi c_p\frac{n}{C}\right)X\left(f-\frac{nB}{C}\right) \\ &= \sum_{n=0}^{C-1}\exp\left(j2\pi c_p\left(\frac{1}{2}-\frac{n}{C}\right)\right)X\left(f-\frac{B}{2}+\frac{nB}{C}\right) \end{aligned} \tag{5.5}$$

where $Y_{c_p}(f) := \frac{C}{B}\cdot X_{c_p}\left(e^{j2\pi f\frac{C}{B}}\right)\exp\left(j2\pi c_p\frac{f}{B}\right)$ for $f \in \left[0, \frac{B}{C}\right)$. The equation (5.5) can be arranged in the matrix form where a CS model with sparse channel occupancy is

revealed,

$$\begin{aligned}\begin{bmatrix} Y_{c_1}(f) \\ Y_{c_2}(f) \\ \vdots \\ Y_{c_P}(f) \end{bmatrix} &= \frac{C}{B}\,\mathrm{diag}_{p=1}^{P}\left(e^{\frac{j2\pi c_p f}{B}}\right)\begin{bmatrix} X_{c_1}\left(e^{j2\pi f\frac{C}{B}}\right) \\ X_{c_2}\left(e^{j2\pi f\frac{C}{B}}\right) \\ \vdots \\ X_{c_P}\left(e^{j2\pi f\frac{C}{B}}\right) \end{bmatrix} \\ &= \boldsymbol{A}\begin{bmatrix} X\left(f-\frac{B}{2}\right) \\ X\left(f-\frac{B}{2}+\frac{B}{C}\right) \\ \vdots \\ X\left(f+\frac{B}{2}-\frac{B}{C}\right) \end{bmatrix},\end{aligned} \tag{5.6}$$

which is then rewritten by vector notations

$$\boldsymbol{y}(f) = \frac{C}{B}\boldsymbol{\Delta}(f)\boldsymbol{x}_c\left(e^{j2\pi f\frac{C}{B}}\right) = \boldsymbol{A}\boldsymbol{x}(f). \tag{5.7}$$

The entries in $\boldsymbol{A}_{P\times C}$ are expressed by $\{\boldsymbol{A}\}_{p,q} = \exp\left(j2\pi c_p(\frac{1}{2}-\frac{q-1}{C})\right)$, for $1\le p\le P$, $1\le q\le C$, where q represents the index of channel. The matrix $\boldsymbol{\Delta}(f) := \mathrm{diag}_{p=1}^{P}\left(e^{\frac{j2\pi c_p f}{B}}\right)$ denotes a diagonal matrix orderly constructed by arguments parameterized by p from 1 to P. In (5.6), it is noted that spectrum sliced into channels, i.e. $\boldsymbol{x}(f)$, has infinite frequency resolution, which leads to the CS linear system (5.6) being an infinite-measurement-vector (IMV) model. The reliable recovery of $\boldsymbol{x}(f)$ is desired given the assumption that the wideband spectrum has sparse occupancy.

In practice, only a finite-length samples can be obtained by each coset, i.e. $\tilde{s_{c_p}}[n] = s_{c_p}[n]\cdot w_p[n]$ for $n = 0, 1, \cdots, N-1$ where N is the number of samples for each coset per sensing round and $w_p[n]$ for $n = 0, 1, \cdots, N-1$ denotes the window function of each coset. Denote the N-point discrete Fourier transform (DFT) of $\tilde{s_{c_p}}[n]$ by $\tilde{X_{c_p}}[k] := \sum_{n=0}^{N-1}\tilde{s_{c_p}}[n]\exp\left(-j2\pi nk/N\right)$. Similarly, the DFT of each coset's samples can be linked to the frequency-domain presentation $\tilde{X}(f)$ of $\tilde{s}(t)$, which is a windowed version of $s(t)$

defined by $\tilde{s}(t) := s(t)w(t)$, writing

$$\begin{aligned} \tilde{X_{c_p}}[k] =& \frac{B}{C} \exp\left(-j2\pi c_p \frac{f}{B}\right) \\ & \cdot \sum_{n=-\infty}^{+\infty} \exp\left(j2\pi c_p \frac{n}{C}\right) \tilde{X}\left(f - \frac{nB}{C}\right)\Big|_{f=\frac{kB}{NC}} \end{aligned} \tag{5.8}$$

where $w(t)$ is the continuous-time-domain presentation of the primitive window function. Discrete-time window function $w_p[n]$ is related to the primitive continuous-time window function $w(t)$ by

$$w_p[n] = w\left(\frac{nC - c_p}{B}\right), \quad n = 0, 1, \cdots, N-1 \tag{5.9}$$

Select the frequency-domain indices $k = 0, 1, \cdots, N-1$ according to the interested period $f \in \left[0, \frac{B}{C}\ (\text{Hz})\right)$ and the relationship $f = \frac{kB}{NC}$ (Hz) from (5.8), and (5.8) is reorganized as

$$\begin{aligned} \tilde{Y_{c_p}}[k] &= \sum_{n=-\frac{C}{2}+1}^{\frac{C}{2}} \exp\left(j2\pi c_p \frac{n}{C}\right) \tilde{X}\left(\frac{kB}{NC} - \frac{nB}{C}\right) \\ &= \sum_{n=0}^{C-1} \exp\left(j2\pi c_p \left(\frac{1}{2} - \frac{n}{C}\right)\right) \tilde{X}\left(\frac{kB}{NC} - \frac{B}{2} + \frac{nB}{C}\right), \end{aligned} \tag{5.10}$$

where $\tilde{Y_{c_p}}[k] := \frac{C}{B} \cdot \tilde{X_{c_p}}[k] \exp\left(j2\pi c_p \frac{k}{NC}\right)$ for $k = 0, 1, \cdots, N-1$. A SMV-model expres-

sion can be directly written for each $k = 0, 1, \cdots, N-1$ based on (5.10),

$$\begin{aligned}\begin{bmatrix} \tilde{Y_{c_1}}[k] \\ \tilde{Y_{c_2}}[k] \\ \vdots \\ \tilde{Y_{c_P}}[k] \end{bmatrix} &= \frac{C}{B} \cdot \mathrm{diag}_{p=1}^{P} \left(e^{\frac{j2\pi c_p k}{NC}} \right) \begin{bmatrix} \tilde{X_{c_1}}[k] \\ \tilde{X_{c_2}}[k] \\ \vdots \\ \tilde{X_{c_P}}[k] \end{bmatrix} \\ &= \boldsymbol{A} \begin{bmatrix} \tilde{X}\left(\frac{kB}{NC} - \frac{B}{2}\right) \\ \tilde{X}\left(\frac{kB}{NC} - \frac{B}{2} + \frac{B}{C}\right) \\ \vdots \\ \tilde{X}\left(\frac{kB}{NC} + \frac{B}{2} - \frac{B}{C}\right) \end{bmatrix} \end{aligned} \tag{5.11}$$

which is then rewritten by vector notation,

$$\tilde{\boldsymbol{y}}[k] = \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}[k] \tilde{\boldsymbol{x}_c}[k] = \boldsymbol{A} \tilde{\boldsymbol{x}}[k], \tag{5.12}$$

where $\tilde{\boldsymbol{\Delta}}[k] := \mathrm{diag}_{p=1}^{P} \left(e^{\frac{j2\pi c_p k}{NC}} \right)$.

If one stacks the N column vectors $\tilde{\boldsymbol{x}_c}[k]$ and $\tilde{\boldsymbol{y}}[k]$ into matrices $\tilde{\boldsymbol{X}}_{C \times N}$ and $\tilde{\boldsymbol{Y}}_{P \times N}$ respectively, the linear system (5.11) can be expressed by a MMV model,

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A} \tilde{\boldsymbol{X}}, \tag{5.13}$$

where, specifically

$$\tilde{\boldsymbol{Y}} := \begin{bmatrix} \tilde{Y_{c_1}}[0] & \tilde{Y_{c_1}}[1] & \cdots & \tilde{Y_{c_1}}[N-1] \\ \tilde{Y_{c_2}}[0] & \tilde{Y_{c_2}}[1] & \cdots & \tilde{Y_{c_2}}[N-1] \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{Y_{c_P}}[0] & \tilde{Y_{c_P}}[1] & \cdots & \tilde{Y_{c_P}}[N-1] \end{bmatrix}$$

and

$$\tilde{\boldsymbol{X}} := \begin{bmatrix} \tilde{X}\left(-\frac{B}{2}\right) & \tilde{X}\left(\frac{B}{NC}-\frac{B}{2}\right) & \cdots & \tilde{X}\left(\frac{(N-1)B}{NC}-\frac{B}{2}\right) \\ \tilde{X}\left(-\frac{B}{2}+\frac{B}{C}\right) & \tilde{X}\left(\frac{B}{NC}-\frac{B}{2}+\frac{B}{C}\right) & \cdots & \tilde{X}\left(\frac{(N-1)B}{NC}-\frac{B}{2}+\frac{B}{C}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{X}\left(\frac{B}{2}\right) & \tilde{X}\left(\frac{B}{NC}+\frac{B}{2}\right) & \cdots & \tilde{X}\left(\frac{(N-1)B}{NC}+\frac{B}{2}\right) \end{bmatrix}.$$

The baseband processing architecture of the multicoset sampler is illustrated in Fig. 5.2.

The signal of interest $\tilde{\boldsymbol{X}}$ is assumed to be row sparse, as a direct result of sparse spectrum usage assumption. It is by definition a discrete 'sliced' spectrum of the windowed baseband signal with the frequency resolution of $\frac{B}{NC}$ (Hz). Recovering a row-sparse matrix $\tilde{\boldsymbol{X}}$ from compressed measurements $\tilde{\boldsymbol{Y}}$ under the sensing matrix $\boldsymbol{A}$ is a fundamental problem in the CS literature [55]. Specifically, for the multicoset samplers architecture in Fig. 5.2, one should set c_p's as distinct integers, i.e. $c_p \in \{0, 1, \cdots, C-1\}$, and $c_p \neq c_r$ for $p \neq r$, making the matrix $\boldsymbol{A}$ a permuted partial Fourier basis, with which good recovery performance is intuitively indicated by its low mutual coherence among all randomly structured dictionary matrices [56]. In the general compressive sensing context, $\boldsymbol{A}$ (after normalization for each row) satisfies the restricted isometric property with high probability hence guarantees reliable recovery if the number of measurements P is large compared to the sparsity of $\boldsymbol{x}(f)$ [122]. Moreover, a necessary condition is given in [65, 125] based on the Kruskal rank of the sensing matrix, that is, arbitrary sparse signal with the sparsity less than $P/2$ can be guaranteed to be recovered with $\boldsymbol{A}$ which has full Kruskal rank (i.e. $\boldsymbol{A}$ is a universal sampling pattern). Therefore, a set of fixed delays c_p's which leads to the sensing matrix $\boldsymbol{A}$ to satisfy RIP or be a universal sampling pattern may be used in practice to provide satisfactory recovery performance. Also, the matrix $\boldsymbol{A}$ is a universal sampling pattern with high probability, if the delays c_p's are drawn uniformly randomly from the set $\{0, 1, \cdots, C-1\}$ [65, 126].

5.2 Spectrum Leakage and Windowing in Multicoset Sampler

It is a well-known fact that any perfectly bandlimited signal has an infinite length in the time-domain. However, in line with requirements of rapid spectrum acquisition in spectrum sensing applications and considering the causality of the practical signal samples, one can only obtain a frame of the infinite signal samples, necessarily causing spectrum leakage. This problem makes the spectrum assumed to be sparse not perfectly sparse. CS algorithms can generally deal with imperfect sparse models (e.g. log-decaying signals) and still exhibit satisfactory recovery performance, especially when the sample length is not too short and the leakage is buried in the additive noise. As a result, such effects are not considered in the literature. However, it is still worth modeling and investigating it as it is found by experiments to impact the performance of the sparsity estimation scheme, and also the CS recovery itself for short sample sizes. Hereby, taking the model of the multicoset sampler, the spectrum leakage problem is revealed in the context of CSS and its impact is proposed to be mitigated by windowing the samples to transform the spectrum leakage into noise-like interference uniformly spreading the spectrum.

By simply considering an N-sample raw frames from the branches of the multicoset sampler, one equivalently applies an N-sample rectangular window

$$w_{\text{rect}}(t) = \begin{cases} 1, \ 0 \leq t < NC/B \\ 0, \ \text{otherwise} \end{cases} \tag{5.14}$$

to each branch. Recall that the Fourier transform of the signal $\tilde{s}(t) := s(t)w_{\text{rect}}(t)$ presents in (5.8), which is then expressed by

$$\begin{aligned} \tilde{X}(f) &= \mathcal{F}\left\{s(t)w_{\text{rect}}(t)\right\} = \mathcal{F}\left\{s(t)\right\} * \mathcal{F}\left\{w_{\text{rect}}(t)\right\} \\ &= X(f) * \left[\frac{B}{NC}\text{sinc}(\frac{B}{NC}f)\exp\left(-j\pi f\frac{B}{NC}\right)\right], \end{aligned} \tag{5.15}$$

where $\mathcal{F}\{\cdot\}$ denotes Fourier transform. From (5.15), it is seen that the spectrum $\tilde{X}(f)$ appearing in $\tilde{\boldsymbol{X}}$ is the original spectrum $X(f)$ convolved with the frequency response of a rectangular window. Consequently, spectrum leakage is present in $\tilde{X}(f)$, specifically causing the energy in the occupied channels to leak to other channels, especially to their adjacent channels. This effect will cause an imperfect row-sparse structure of $\tilde{\boldsymbol{X}}$, which may impose negative effects on the sparsity estimation and the row sparse recovery procedures.

To address the problem of spectrum leakage, it is proposed to use an NC-sample Dolph-Chebyshev window function, $w_{\text{dch}}[m], \quad 0 \leq m \leq CN-1$, on top of the raw samples. The frequency response of a Dolph-Cheybyshev window has equiripple sidelobe with minimum magnitude given mainlobe bandwidth. This means that spectrum leakage can be controlled such that the majority of the leakage power is concentrated in the mainlobe bandwidth around the channel edges, while a small amount of power evenly leaks into the rest of the spectrum. The multicoset sampling process with the proposed windowing is then expressed by

$$\tilde{s_{c_p}}[n] = w_{\text{dch}}\left[nC - c_p\right] \cdot s\left(\frac{nC - c_p}{B}\right), \quad 0 \leq n \leq N-1. \tag{5.16}$$

Accordingly, the expression of the spectrum $\tilde{X}(f)$ in (11) changes to

$$\begin{aligned} \tilde{X}(f) &= X(f) * \mathcal{F}\left\{\sum_{n=0}^{NC-1} w_{\text{dch}}[n]\delta(t + \frac{nC}{B})\right\} \\ &= X(f) * \left[\sum_{n=0}^{NC-1} w_{\text{dch}}[n] \exp\left(j2\pi\frac{nC}{B}\right)\right] \end{aligned} \tag{5.17}$$

Note that the last term in (5.17) is actually the DTFT of the chosen NC-sample Dolph-Chebyshev window, which has the following closed-form expression derived from Chebyshev polynomial [195],

$$W_{\text{dch}}(f) = \frac{\cos\left\{NC\cos^{-1}\left[b\cos\left(\frac{\pi fC}{B}\right)\right]\right\}}{\cosh\left[NC\cosh^{-1}(b)\right]}, \tag{5.18}$$

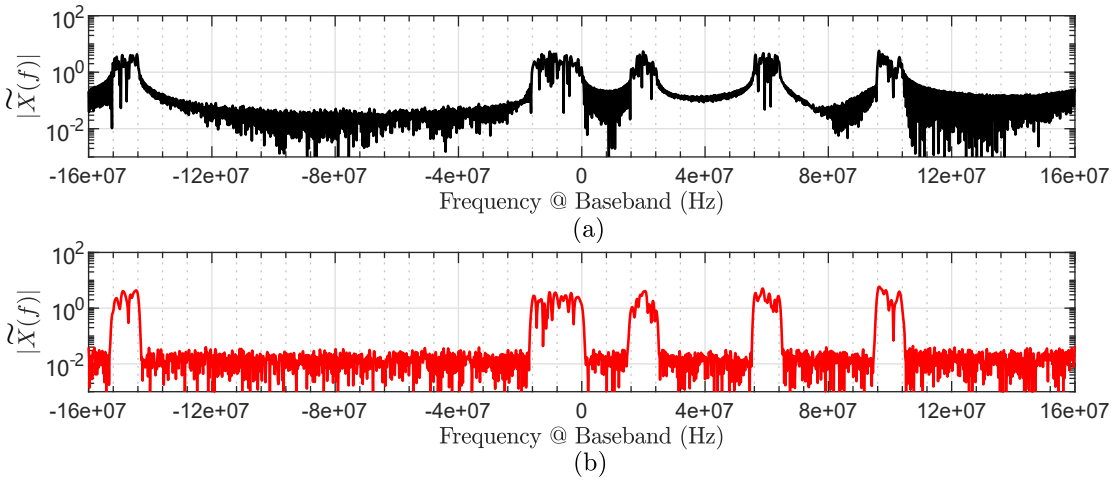


Figure 5.3: Presentation of the amplitude of $\tilde{X(f)}$ **(a)** without the proposed windowing as in (5.15); **(b)** with Dolph-Chebyshev windowing as in (5.17).

where $b := \cosh\left[\frac{1}{NC}\cosh^{-1}(\frac{1}{a})\right]$ is the parameter related to the ripple magnitude a. The mainlobe edge frequency can be then found related to a [195],

$$f_{\text{dch}} = \frac{B}{\pi C}\cos^{-1}\left(\frac{1}{b}\right) \text{ (Hz)}. \tag{5.19}$$

An example of comparison between $\tilde{X(f)}$ in (5.15) (i.e. without the proposed windowing) and in (5.17) (i.e. with Dolph-Chebyshev windowing) with $a = 10^{-2}$ is presented in Fig. 5.3.

Since most of the spectrum leakage appears in the half mainlobe bandwidth of the Dolph-Chebyshev window around the channel edges, $f = -\frac{B}{2}, -\frac{B}{2} + \frac{B}{C}, \cdots, \frac{B}{2}$, in order to minimize the impact of spectrum leakage, it is proposed to remove the first and last W columns of which the corresponding frequencies falling into the mainlobe bandwidth in the sliced spectrum $\tilde{\boldsymbol{X}}$. W is determined by

$$W = \lceil f_{\text{dch}} \rceil = \left\lceil \frac{1}{\pi}\cos^{-1}\left(\frac{1}{b}\right) \right\rceil, \tag{5.20}$$

where $\lceil \cdot \rceil$ represents ceiling operation.

As a result of the removal of $2W$ columns, the MMV linear relationship in (5.13) can

be updated by

$$\hat{\boldsymbol{Y}}_{P\times(N-2W)} = \boldsymbol{A}_{P\times C}\hat{\boldsymbol{X}}_{C\times(N-2W)} + \hat{\boldsymbol{B}}, \tag{5.21}$$

where

$$\hat{\boldsymbol{Y}} = \left[\left(\tilde{\boldsymbol{Y}}\right)_{\cdot,W+1}, \left(\tilde{\boldsymbol{Y}}\right)_{\cdot,W+2}, \cdots, \left(\tilde{\boldsymbol{Y}}\right)_{\cdot,N-2W}\right]$$

,

$$\hat{\boldsymbol{X}} = \left[\left(\tilde{\boldsymbol{X}}\right)_{\cdot,W+1}, \left(\tilde{\boldsymbol{X}}\right)_{\cdot,W+2}, \cdots, \left(\tilde{\boldsymbol{X}}\right)_{\cdot,N-2W}\right],$$

and $\hat{\boldsymbol{B}}$ is additive noise. Due to the equi-ripple property of the Dolph-Chebyshev window's sidelobe, the spectrum leakage after removing $2W$ column in (5.21) can be effectively accounted by the noise term $\hat{\boldsymbol{B}}$, and one can consider

$$\operatorname{supp}\left(\hat{\boldsymbol{X}}\right) = \operatorname{supp}\left(\boldsymbol{X}\right). \tag{5.22}$$

5.3 Row Sparsity Estimation and Model Dimension Reduction for Multicoset Sampler

To enable the row-sparse signal $\hat{\boldsymbol{X}}$ in the MMV model (5.21) of the multicoset sampler to be solved by greedy algorithms, an estimate of row sparsity has to be provided to the algorithms. Hereby, the sparsity estimation scheme proposed in Section 3.3 for the SMV problem is extended to the MMV setting to be useful for multicoset sampler.

With a finite number L of measurement frames, an observation set is obtained, denoted by $\widehat{\boldsymbol{Y}_{\text{obs}}} := \left[\hat{\boldsymbol{Y}}^{(1)}, \hat{\boldsymbol{Y}}^{(2)}, \cdots, \hat{\boldsymbol{Y}}^{(L)}\right]$. A practical consistent estimation of the autocorrelation of $\hat{\boldsymbol{Y}}$, i.e. $\mathrm{E}\left[\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}^H\right]$, denoted by $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}} = \frac{1}{L}\sum_{l=1}^{L}\hat{\boldsymbol{Y}}^{(l)}\left(\hat{\boldsymbol{Y}}^{(l)}\right)^H$, satisfies

$$\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}} = \left(\frac{B}{C}\right)^2 \boldsymbol{A}\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{X}}}}\boldsymbol{A}^H + \frac{1}{L}\sum_{l=1}^{L}\hat{\boldsymbol{B}}^{(l)}\left(\hat{\boldsymbol{B}}^{(l)}\right)^H, \tag{5.23}$$

where $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{X}}}} = \frac{1}{L}\sum_{l=1}^{L}\hat{\boldsymbol{X}}^{(l)}\left(\hat{\boldsymbol{X}}^{(l)}\right)^H$ and elements in $\hat{\boldsymbol{B}}^{(l)}$ are assumed to be independent

random variables and each follows $\mathcal{CN}(0, \sigma^2)$. Assuming the spectrum occupancy is static during the observation of L frames, one can arrive at a proposition which guarantees the rank of $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}$ to be determined by the row-sparsity of $\hat{\boldsymbol{X}}$ in the noiseless setting.

Proposition 5. *Given $\hat{\boldsymbol{B}} = \boldsymbol{0}$, and that random matrix $\hat{\boldsymbol{X}}$ has the row-sparsity of $\delta(\hat{\boldsymbol{X}})$, if $(N - 2W)L \geq \delta(\hat{\boldsymbol{X}})$, it holds with probability of 1 that rank$(\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}) =$ rank$(\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{X}}}}) = \delta(\hat{\boldsymbol{X}})$.*

Proof. See [54, Lemma 1, Proposition 1]. □

The conclusion in Proposition 1 implies that in the noiseless setting one could perform eigendecomposition and determine the row-sparsity by counting the number of non-zero eigenvalues. Express the eigendecomposition of the estimated autocorrelation of the true measurement matrices

$$\left(\frac{B}{C}\right)^2 \boldsymbol{A}\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{X}}}}\boldsymbol{A}^H = \sum_{i=1}^{\delta(\hat{\boldsymbol{X}})} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^H, \tag{5.24}$$

where λ_i and $\boldsymbol{v}_i$ are the ith eigenvalue and corresponding eigenvector, $1 \leq i \leq \delta(\hat{\boldsymbol{X}})$.

In the presence of noise, the following is obtained from (5.23) and (5.24),

$$\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}} = \sum_{i=1}^{\delta(\hat{\boldsymbol{X}})} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^H + \frac{1}{L}\sum_{l=1}^{L} \hat{\boldsymbol{B}}^{(l)} \left(\hat{\boldsymbol{B}}^{(l)}\right)^H = \sum_{i=1}^{P} \hat{\lambda}_i \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_i{}^H, \tag{5.25}$$

where $\hat{\lambda}_i$ and $\hat{\boldsymbol{v}}_i$ denote the ith largest eigenvalue and corresponding eigenvector of noisy $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}$, $1 \leq i \leq P$. With relatively large signal-to-noise ratio (SNR), the noise will perturb the $\delta(\hat{\boldsymbol{X}})$-dimensional signal space spanned by $\boldsymbol{v}_i$'s, and the resultant eigenvalues and eigenvectors in the noisy scenario satisfy $\hat{\lambda}_i \approx \lambda_i$ and $\hat{\boldsymbol{v}}_i \approx \boldsymbol{v}_i$, $1 \leq i \leq \delta(\hat{\boldsymbol{X}})$, as a result of high SNR. The eigenspaces spanned by $\hat{\boldsymbol{v}}_{\delta(\hat{\boldsymbol{X}})+1}, \hat{\boldsymbol{v}}_{\delta(\hat{\boldsymbol{X}})+2}, \cdots, \hat{\boldsymbol{v}}_P$ is determined by noise only, whose amplitudes are small compared to those of signal subspace and asymptotically converge to $\hat{\boldsymbol{v}}_{\delta(\hat{\boldsymbol{X}})+1} = \hat{\boldsymbol{v}}_{\delta(\hat{\boldsymbol{X}}+2)} = \cdots = \hat{\boldsymbol{v}}_P = \sigma^2$ with the increasing number of frames L.

The sparsity estimation problem then translates into finding the integer p^o $(1 \leq$

$p^o < P$) splitting the P descending-ordered eigenvalues into two groups, corresponding to signal subspace and noise subspace. Formally, this can be achieved by minimizing some information theoretic criterion (ITC). The enhanced Bayesian information criterion (BICe) [163] is hereby proposed to be employed, whose objective function is effectively the joint probability of all the observations, $\widetilde{\boldsymbol{Y}_{\text{obs}}}$, and signal subspace eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$. The objective of the proposed BICe-based row-sparsity estimation writes

$$\begin{aligned} p^o &= \arg\min_p \text{BICe}(p) \\ &= \arg\min_p \big\{ -2\log f(\widetilde{\boldsymbol{Y}_{\text{obs}}}|\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_P; p) \\ &\quad -2\log f(\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_P; p) - C_p \log\left[(N-2W)L\right]\big\}, \end{aligned} \tag{5.26}$$

where $f(\cdot)$ denotes probability density function and $C_p = p(2P-p)$ is the number of free parameters. Following the statistical models in [54, 167], one can rearrange the estimation objective function as follows,

$$\begin{aligned} p^o &= \arg\min_p \text{BICe}(p) \\ &= \arg\min_p \big\{ 2(P-p)\left[(N-2W)L + P - p\right]\log\hat{\sigma^2} \\ &\quad -2(N-2W)L\sum_{i=p+1}^{P}\log\hat{\lambda}_i + 2\sum_{i=1}^{p}\log\hat{\lambda}_i + 2\sum_{i=1}^{P-p}\log\Gamma(i) \\ &\quad +(N-2W)L\sum_{i=p+1}^{P}\left(\frac{\hat{\lambda}_i - \hat{\sigma^2}}{\hat{\sigma^2}}\right)^2 - 4\sum_{i=p+1}^{P}\sum_{k=i+1}^{M}\log|\hat{\lambda}_i - \hat{\lambda}_k| \\ &\quad +p(4P-2p-1)\log\left[(N-2W)L\right]\big\}, \end{aligned} \tag{5.27}$$

where $\hat{\sigma^2} := \sum_{i=p+1}^{P}\hat{\lambda}_i/(P-p)$ is an estimate of noise variance and $\Gamma(\cdot)$ denotes Gamma function.

Having a valid estimation of the row-sparsity p^o of the sliced spectrum $\hat{\boldsymbol{X}}$, one can reduce the dimension of the MMV linear system (5.21) from $N-2W$ to p^o by removing the components in the noise subspace. Denote the $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}$'s component in the signal subspace by $\sum_{i=1}^{p^o}\hat{\lambda}_i\hat{\boldsymbol{v}}_i\hat{\boldsymbol{v}}_i{}^H = \boldsymbol{V}_s\boldsymbol{\Lambda}_s\boldsymbol{V}_s^H$, where $(\boldsymbol{V}_s)_{P\times p^o} = [\hat{\boldsymbol{v}_1}, \hat{\boldsymbol{v}_2}, \cdots, \hat{\boldsymbol{v}_{p^o}}]$ and

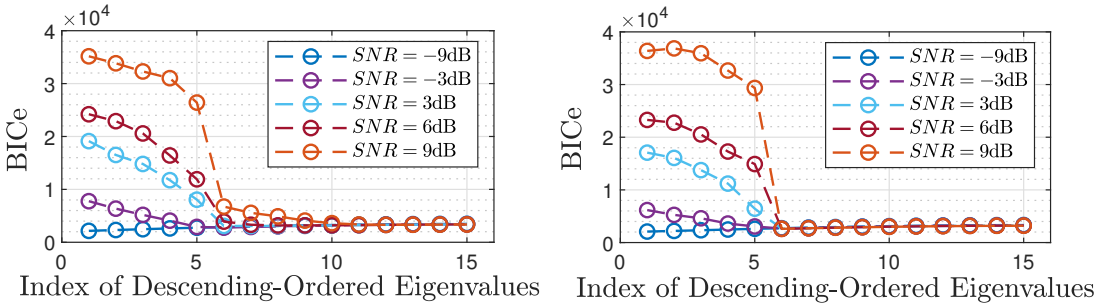


Figure 5.4: BICe profiles of $\widetilde{\boldsymbol{R}_{\hat{\boldsymbol{Y}}}}$ **(a)** with no windowing; **(b)** with Dolph-Chebyshev windowing.

$(\boldsymbol{\Lambda}_s)_{p^o\times p^o} = \operatorname{diag}\left(\hat{\lambda_1}, \hat{\lambda_2}, \cdots, \hat{\lambda_{p^o}}\right)$. The reduced MMV linear system can be then expressed by

$$\boldsymbol{V}_s \left(\boldsymbol{\Lambda}_s\right)^{\frac{1}{2}} = \boldsymbol{A}\hat{\boldsymbol{X}}_v + \hat{\boldsymbol{B}}_v, \tag{5.28}$$

where $\hat{\boldsymbol{B}}_v$ denotes the noise perturbing the signal subspace, and it is guaranteed [16, 54] that $\operatorname{supp}\left(\left(\hat{\boldsymbol{X}}_v\right)_{C\times p^o}\right) = \operatorname{supp}\left(\hat{\boldsymbol{X}}\right)$. Furthermore, from (5.22), one also gets the support of the original signal $\operatorname{supp}\left(\boldsymbol{X}\right) = \operatorname{supp}\left(\hat{\boldsymbol{X}}_v\right)$.

After obtaining the support of spectrum, according to the original form of the MMV model (5.13), it takes a single step of projection to recover $\tilde{\boldsymbol{X}}$ itself

$$\tilde{\boldsymbol{X}} \leftarrow \left(\boldsymbol{A}_{\operatorname{supp}(\boldsymbol{X})}\right)^{\dagger}\tilde{\boldsymbol{Y}} \tag{5.29}$$

where $(\cdot)^{\dagger}$ denotes psuedo-inverse and $\boldsymbol{A}_{\operatorname{supp}(\boldsymbol{X})}$ is the submatrix of $\boldsymbol{A}$ whose columns are indexed by $\operatorname{supp}(\boldsymbol{X})$.

The performance of the proposed BICe-based row-sparsity estimator is analyzed. The baseband frequency range is set as $[-160\ (\text{MHz}), 160\ (\text{MHz})]$ consisting of 40 channels, making $B = 320\ (\text{MHz})$ and $C = 40$. The length of frame is set to $N = 32$ and the number of frames in the observation set is $L = 16$ unless otherwise stated. 6 out of 40 channels are set occupied. The number of cosets is fixed at $P = 16$. Time-domain

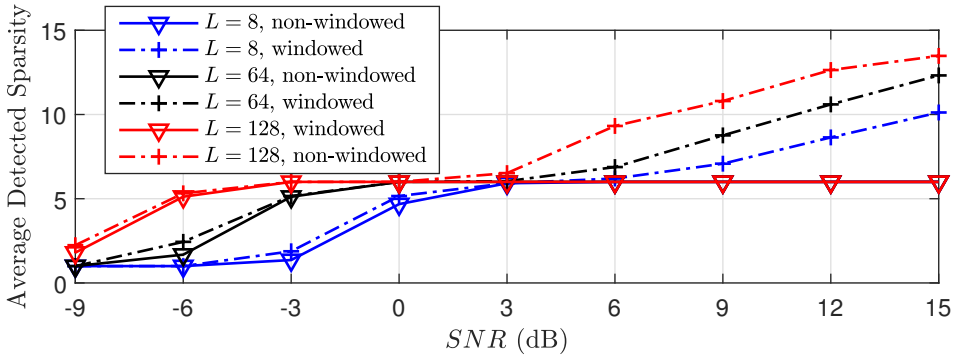


Figure 5.5: The average detected row-sparsity with and without the proposed windowing over 1000 simulations, with varying SNR and number of frames L.

OFDM signals $s(t)$ are generated by

$$s(t) = \sum_{c=1}^{C} \sum_{d=1}^{\frac{B}{CF_s}} \eta_{c,d} e^{j2\pi[(d-1)F_s + (c-1)B]t}, \tag{5.30}$$

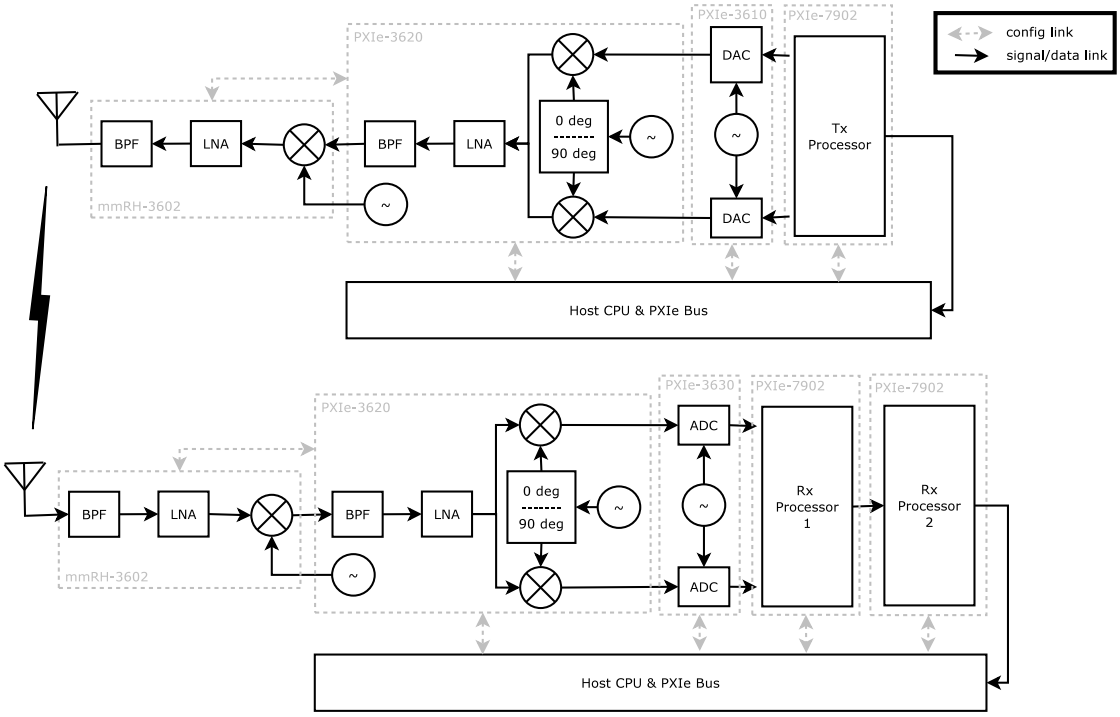
where $\eta_{c,d}$ is complex Gaussian-distributed symbols and $F_s = 39.0625$ (kHz) is the OFDM symbol rate. SNR is formally defined by $SNR := ||\tilde{\boldsymbol{X}}||_F^2/(\sigma^2 NC)$. A 1280-sample Dolph-Chebyshev window is used with $a = 10^{-3}$ and $M = 2$ is determined from (5.20). In Fig. 5.4, the BICe profiles of the estimated autocorrelation matrix $\widetilde{\boldsymbol{R}_{\tilde{\boldsymbol{Y}}}}$ in (5.21) is directly given with comparison with the estimated autocorrelation $(1/L)\sum_{l=1}^{L} \tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H$ where $\tilde{\boldsymbol{Y}}$ corresponds to sampling without the proposed windowing (i.e. with rectangular windowing). The impact of spectrum leakage mentioned in Section 5.2 is apparent - the BICe profile without the proposed windowing tend to have its minimum at indices over the spectrum's true row-sparsity, i.e. 6, while the BICe's minimal with Dolph-Chebyshev windowing converges to 6 with increasing SNR. In Fig. 5.5, the row-sparsity estimation performance with and without the proposed windowing are illustrated against varying SNR and frame count L. As an important trend, the estimation performance has more robustness against noise with a larger observation set of frames. Moreover, the spectrum leakage causes severe overestimation, especially when noise is not large enough to cover spectrum leakage, making the estimator unusable. Hence, the proposed Dolph-Chebyshev windowing is necessary for robust row-sparsity estimation, although it

has very slight performance deterioration under low SNRs as a result of removing $2W$ columns in (5.21).

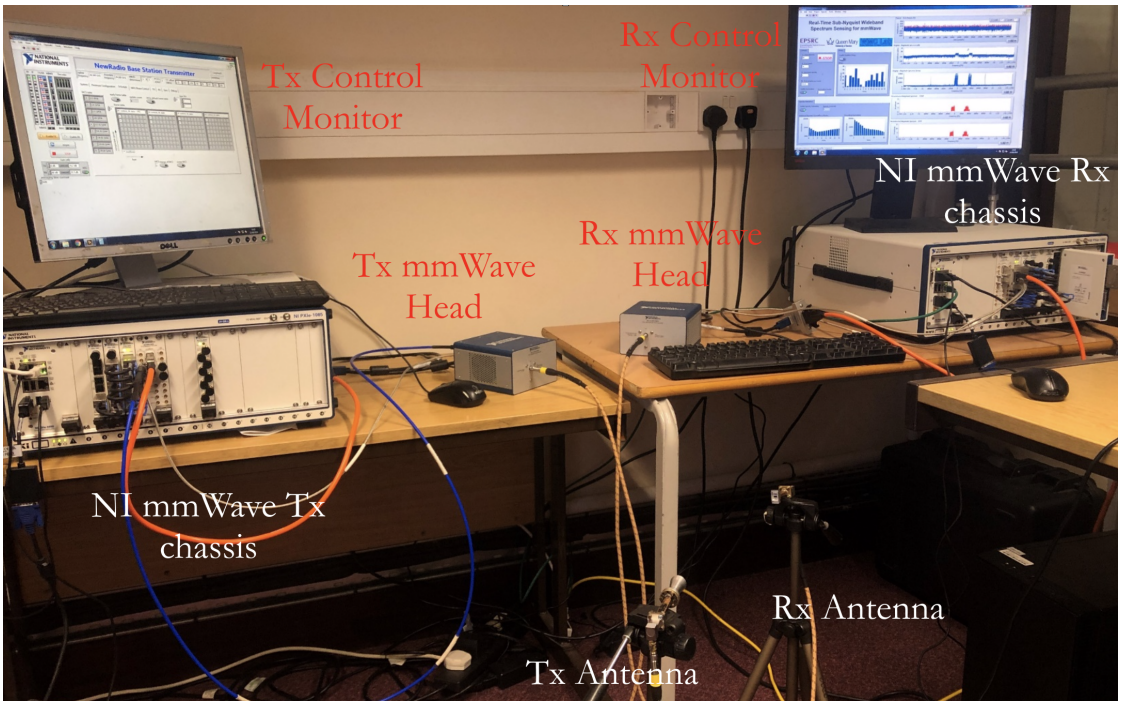
5.4 Software-Defined-Radio Demo for mmWave-Band Multicoset-Sampler-Based Compressive Spectrum Sensing

Having detailed the practical model and a few technical considerations of the multicoset sampler in this chapter, it is of great interest to implement such the multicoset-sampler-based compressive spectrum sensing (CSS) in a real-world wideband receiver system. The work on realizing a real-time sub-Nyquist CSS system demo based on NI mmWave SDR platform [196] is hereby demonstrated. In this experiment, two sets of the NI mmWave SDR devices are used, as the Tx and RX respectively, as shown in Fig. 5.6 (a). Each device features a host computer with a peripheral component interface express (PCIe) bus and the full superheterodyne transmitter or receiver architecture, consisting of a front-end module (namely mmWave head) and a few pieces of modular PCIe cards as the all necessary components in the Tx or Rx chain. Those Tx/Rx chain modules can be controlled and programmed by software on the host computer through the PCIe bus. Baseband processing can be performed by one or more hard processors PXIe-7902, each of which features a Xilinx Virtex-7 FPGA [197]. Moreover, those processing tasks not suitable to be implemented in FPGA and intermediate data can be offloaded to the host computer for further processing and input/output (I/O) purposes. The setup appearance of the two SDR devices are shown in Fig. 5.6 (b).

At the transmitter side, multiple orthogonal frequency division access (OFDM) signals conforming to the Verizon 5G New Radio physical layer specifications are generated, each having a 3.072 GHz symbol rate spanning the bandwidth of 100 MHz. Such multi-band frequency-sparse signals are upconverted to the 12 GHz intermediate frequency (IF), and then to the mmWave radio frequency (RF) of 28.5 GHz. At the receiver end, reversely, such RF signals are downconverted from the RF central frequency to IF and then to the baseband. The baseband in-phase (I) and quadrature (Q) signals are



(a)



(b)

Figure 5.6: **(a)** System architecture of National Instrument mmWave SDR platform. **(b)** Pictorial presentation of the TX and RX SDR setup.

respectively sampled at 3.072 GHz [196].

To achieve the interleaved sub-Nyquist sampling pattern of the multicoset sampler, only part of the 3.072 GHz (Nyquist-rate) samples are retained for processing. As indicated in Fig. 5.2, measurements $\tilde{X_{c_p}}[k]$ ($0 \leq k \leq N-1$ and $1 \leq p \leq P$) are obtained through a parallel structure and the pre-recovery processing is implemented in one of the FPGA card. The measurements are then transferred from the FPGA card to the host via the PCIe bus and the real-time spectrum is recovered from the measurements at the host CPU.

Apart from the configuration on the ready-to-use transmitter-side 5G-physical-layer reference design, the main development work is carried out using NI LabView and Labview FPGA toolchains and contains developments of the programs for three parts, with their main functions summarized as below. [1]

- Rx processor 1 (PXIe-7902) - interfacing with the ADC card (PXIe-3630); converting the two (I and Q) 3.072 GHz 16-bit serial data streams from the ADC card into two 192 MHz 256-bit parallel data stream to adapt to FPGAs' timing constraints; transmitting the paralleled data to Rx processor 2 via two 4-channel high-speed serial interfaces for I and Q respectively, each fulfilling the 49.152 Gbps data rate; a routine of calculating the average power of the baseband data to be fed to the host to achieve automatic gain control of the Rx chain;

- Rx processor 2 (PXIe-7902) - interfacing with the Rx processor 1 and receiving two streams of I and Q data; decimating the Nyquist rate samples according to the interleaved sampling pattern of the multicoset sampler; splitting the decimated data into C ($C \leq 16$ due to FPGA resource and timing constraints) branches, each having an N-sample frame; each of the branches performing windowing and N-FFT in real-time as shown in Fig. 5.2; a first-in-first-out (FIFO) buffer at the end of each branch caching and feeding the measurements $\tilde{X_{c_p}}[k]$'s through

[1] Credit to Mr. Zihang Song for his contributions to the software development of this demo.

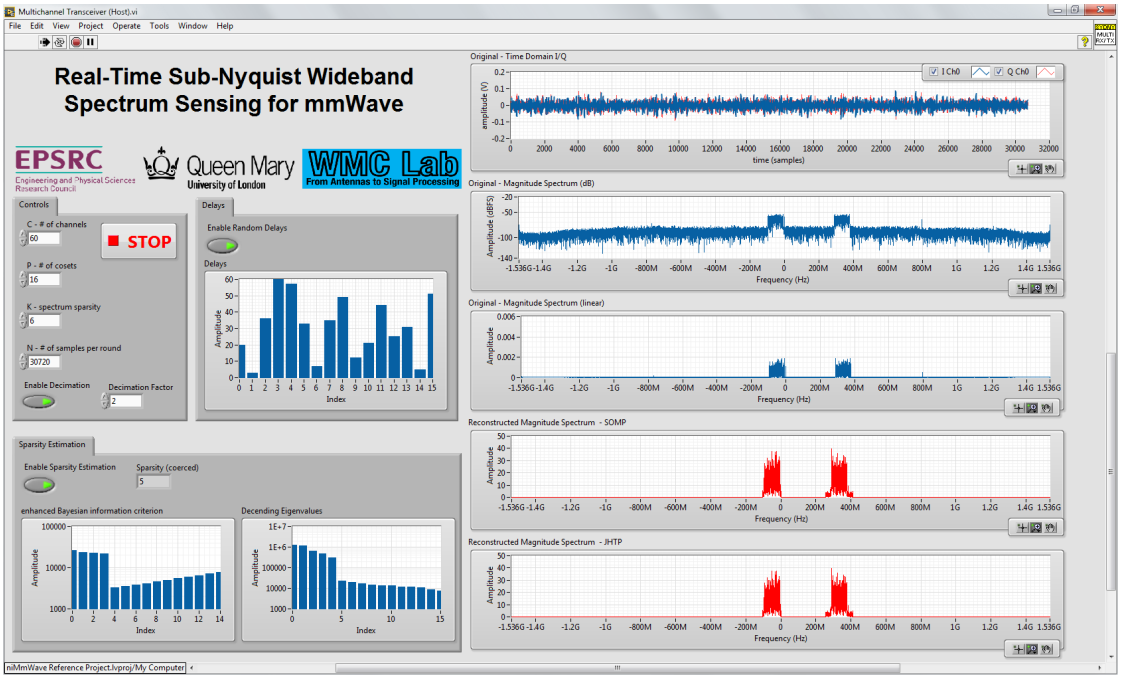


Figure 5.7: Control and display panel of the SDR demo.

the direct-memory-access (DMA) process of the PCIe bus to the host computer's memory.

- Host computer - initializing and configuring the Rx chain, including the mmWave head mmRH-3602, the IF to baseband converter PXIe-3620, and the ADC card PXIe3630; triggering the data streaming between Rx processor 1 and 2, and the FIFO-DMA transfer between the Rx processor 2 and the host, through the PCIe bus; issuing the multicoset sampler pattern c_p's and parameters P, C and N to the Rx processors; reading the DMA memory region and performing the proposed sparsity estimation and dimension reduction as discussed in Section 5.2; applying greedy sparse recovery algorithms on the dimension-reduced MMV problem (5.28) and (5.29) to obtain the real-time reconstructed spectrum $\hat{\boldsymbol{X}}$; providing a graphical user interface (GUI) for configuration and control inputs and real-time spectrum display as shown in Fig. 5.7.

The GUI panel of the host program is as presented in Fig. 5.7. The parameters for the multicoset sampler can be input at the left-hand side. The sampling pattern, i.e. the

timing offsets of the interleaved samples, is generated randomly by the host. The left bottom shows the sparsity estimation results, along with the eigenvalue profiles and the BICe values discussed in Section 5.2. On the right, the original data frames are directly acquired from the Rx processor 1 by DMA and displayed for comparison, in the form of time-domain data, logarithmic and linear scale power spectrum with blue color. Two versions of the recovered spectrum are shown in red color, with SOMP and JB-HTP greedy algorithms (presented in Chapter 3) respectively implemented on the host.

It should be noted that the reason why the greedy sparse recovery is implemented on the host is that the greedy algorithm requires arithmetics of complex and float-point valued matrices, such as inversion, which is considered challenging and unsuitable to be implemented by hardware and that the parallel signal processing routines of the multicoset sampler already requires a large number of hardware resources of the FPGA. Moreover, there are technical challenges as the following related to the throughput of the proposed demo system

- Could the timing constraints of the Rx processor 2 FPGA be met with two 192 MHz 256-bit data streams from the raw ADC output and the hardware utilization of 16 branches of windowing and FFT blocks?
- Could the average running time for the greedy sparse recovery on the host be less than the time of a single frame so that the real-time recovery becomes possible?

These problems are more or less platform-dependent. The first one may be mitigated by using an FPGA model with abundant resources and applying careful timing optimization of the FPGA design. However, the latter is more of a common problem for all kinds of platforms. For instance, the time consumption of greedy algorithms on the NI SDR host computer is empirically of the magnitude order of 10 ms, far exceeding the time of a frame, 10 us. To alleviate such a problem, and also to relieve the timing constraints of the FPGA design, it is necessary to reduce the system throughput to allow for real-time spectrum recovery. Hereby it is proposed to include an additional 'frame summing' stage

right after the 196 MHz data stream input at the Rx processor 2. Define the summing factor D and suppose there are D consecutive frames of N-point time-domain Nyquist rate data

$$s[n] = s(t)|_{t=n/B} \qquad n = 0, 1, \cdots, DN-1, \tag{5.31}$$

where $s(t)$ is the baseband signal and $1/B$ the Nyquist rate. By summing all D frames point by point, one can obtained the summed frame as

$$s_{\mathrm{SUM},D}[n] = \sum_{d=0}^{D-1} s[dN + n - 1], \tag{5.32}$$

without discarding any time-domain sample. Looking at the frequency-domain representation of the Nd-sample data and the averaged N-sample sequence, one can write the DFT coefficients as

$$\begin{aligned} X[k] &= \sum_{n=0}^{DN-1} s[n]\exp\Big(-j2\pi k\frac{n}{DN}\Big) \\ X_{\mathrm{SUM},D}[k] &= \sum_{n=0}^{N-1}\sum_{d=0}^{D-1} s[dN + n - 1]\exp\Big(-j2\pi k\frac{n}{N}\Big) \end{aligned} \tag{5.33}$$

where it is revealed

$$X[Dk] = X_{\mathrm{SUM},D}[k], \qquad k = 0, 1, \cdots, N-1. \tag{5.34}$$

The relationship above implies that the spectrum of the summed frame equivalently consists of the decimated frequency bins of the spectrum of the DN-sample sequence - in the frequency domain every 1 of D consecutive frequency bins are preserved as is. Applying multicoset sampling pattern on the summed frames reduce the data throughput to $\frac{1}{D}$ of the raw data throughput. Recall the original sampling pattern for multicoset sampler is as (5.1), rewritten as follows

$$s_{c_p}[n] = s(t)\big|_{t=\frac{nC-c_p}{B}} = s[nC - c_p]. \tag{5.35}$$

After applying the proposed frame summing procedure, the multicoset sampler works on the summed frames, formally

$$s_{c_p}[n] = s_{\mathrm{SUM},D}[nC - c_p] = \sum_{d=0}^{D-1} s[dN + nC - c_p - 1], \qquad n = 0, 1, \cdots, N-1. \quad (5.36)$$

With this proposed throughput reduction method by summing the consecutive frames, real-time sparse recovery implemented on the host computer is made possible without discarding some of the time-domain frames, thus avoiding invisibility of some short-lived incumbent radio activities. On the other hand, this approach only preserves one out of D original frequency bins, implying compromised spectrum resolution, which could impose a problem for narrowband signals.

5.5 On Timing Skews of Practical Multicoset Samplers

With the motivation to bring the real-time wideband CSS implementation from TV whitespace [15, 16, 161] towards multi-gigahertz bandwidth system in mmWave bands, the hardware requirements of multicoset sampler pose larger challenges in the precision of timing control, and the accuracy of compressed measurements is hence more prone to timing errors. It is noted that in the multicoset sampling pattern (5.1), the resolution of the sample timing offset is $1/B$, i.e. the Nyquist sampling period, and with the significant increase of the baseband instantaneous bandwidth for mmWave applications, precise timing control is essential for such interleaved sampling. For example, if the frequency spectrum being sampled has a bandwidth of 2 GHz, the timing offset step should be around 500 ps, and the error tolerance of the offset should be only a fractional of 500 ps to minimize the error due to imperfect timing in the obtained samples. Proposals for the implementation of the multicoset sampler in the literature have relied on delay filters on the analog paths in front of the ADCs to achieve time-shifted versions of the baseband signal [194, 198, 199]. For multicoset samplers, it is preferred that randomly selected delays are used to form a universal sampling matrix $\boldsymbol{A}$ to ensure guaranteed recovery performance [61]. Therefore, the delays should be reconfigurable in order

to obtain the optimal performance of the signal reconstruction. Current reconfigurable analog delay lines in literature largely rely on transistor switching architecture or tunable line elements, and existing implementations have limited delay reconfigurability and low bandwidth to be applicable in multicoset samplers [200–203]. An alternative and more practical strategy to achieve reconfigurable delays is to apply delays on the sampling clocks instead, by using clock distribution techniques, for example, delay-locked loops and delay-locked loops. In the real-world implementation of the multicoset sampler, regardless of the specific techniques used to achieve the delays, however, the precision of the timing offset of the sampling is always finite [194], and it is of particular interest at the system level to quantify the error in the compressed measurements induced by timing skews, which further leads to performance degradation of signal reconstruction. Although additive white Gaussian noise is widely discussed in most evaluations of multicoset samplers to model the thermal noise of the receiver [15, 16], the analysis of timing skew and its impact in multicoset samplers is still absent in the literature.

The error analysis based on the timing skews of multicoset samplers is presented in Section 5.5.1. Verification of the analytic forms of measurement error is provided in Section 5.5.2, where the impacts of such error to signal recovery are initially evaluated.

5.5.1 Error Analysis for Timing Skews

In this subsection, the amount of error at the measurements caused by timing skews of all cosets is examined. From the implementation perspective in Fig. 5.2, such timing skews can be caused by signal trace length mismatch, design limitations of the delay units, aperture delay discrepancies across ADCs, and skews of the sampling clocks. Start by separating the offset at each coset into two components, i.e.

$$c_p = \tau_p + \delta_p \tag{5.37}$$

where τ_p is the desired offset and δ_p denotes the skew.

5.5.1.1 Error in SMV Measurements

Recall (5.12) and rewrite it as follows

$$\tilde{\boldsymbol{x}_c}[k] = \frac{B}{C} \cdot \left(\tilde{\boldsymbol{\Delta}}[k]\right)^{-1} \boldsymbol{A}\tilde{\boldsymbol{x}}[k], \tag{5.38}$$

where $\tilde{\boldsymbol{\Delta}}[k]$ and $\boldsymbol{A}$ are determined by the values of c_p's. Separate $\tilde{\boldsymbol{\Delta}}[k]$ into $\tilde{\boldsymbol{\Delta}_\tau}[k]$ and $\tilde{\boldsymbol{\Delta}_\delta}[k]$ which only depend on τ_p's and δ_p's respectively, and one may easily obtain the following

$$\tilde{\boldsymbol{\Delta}}[k] = \tilde{\boldsymbol{\Delta}_\tau}[k] \cdot \tilde{\boldsymbol{\Delta}_\delta}[k], \tag{5.39}$$

where $\tilde{\boldsymbol{\Delta}_\tau}[k] = \mathrm{diag}_{p=1}^{P}\left(\left(\exp\left(j2\pi\tau_p\frac{k}{NC}\right)\right)\right.$ and $\tilde{\boldsymbol{\Delta}_\delta}[k] = \mathrm{diag}_{p=1}^{P}\left(\exp\left(j2\pi\delta_p\frac{k}{NC}\right)\right)$. As for matrix $\boldsymbol{A}$, the element-wise presentation rewrites

$$\begin{aligned}\{\boldsymbol{A}\}_{p,q} &= e^{\left(j2\pi c_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)} \\ &= e^{\left(j2\pi\tau_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)} e^{\left(j2\pi\delta_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)},\end{aligned} \tag{5.40}$$

In order to separate $\boldsymbol{A}$ into $\boldsymbol{A}_\tau$ and $\boldsymbol{A}_\delta$, one may linearize the exponential term relevant to δ_p's using Taylor's series for δ_p's near 0, which is

$$e^{\left(j2\pi\delta_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)} = 1 + \sum_{n=1}^{+\infty} \frac{\left(j2\pi\delta_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)^n}{n!}. \tag{5.41}$$

Therefore, one can separate $\boldsymbol{A}$ under addition,

$$\boldsymbol{A} = \boldsymbol{A}_\tau + \boldsymbol{A}_\delta, \tag{5.42}$$

where

$$\{\boldsymbol{A}_\tau\}_{p,q} = e^{\left(j2\pi\tau_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)} \tag{5.43}$$

and

$$\{\boldsymbol{A}_\delta\}_{p,q} = e^{\left(j2\pi\tau_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)} \sum_{n=1}^{+\infty} \frac{\left(j2\pi\delta_p\left(\frac{1}{2}-\frac{q-1}{C}\right)\right)^n}{n!}. \tag{5.44}$$

Insert (5.39) and (5.42) into (5.38), and one has the following

$$\tilde{\boldsymbol{x}_c}[k] = \frac{B}{C} \cdot \left(\tilde{\boldsymbol{\Delta}_\tau}[k]\right)^{-1} \left(\tilde{\boldsymbol{\Delta}_\delta}[k]\right)^{-1} (\boldsymbol{A}_\tau + \boldsymbol{A}_\delta)\, \tilde{\boldsymbol{x}}[k]. \tag{5.45}$$

Suppose the measurements $\tilde{\boldsymbol{x}_c}[k]$ compromises a skew-less component $\tilde{\boldsymbol{x}_\tau}[k]$ only relevant to τ, and an error term $\tilde{\boldsymbol{x}_e}[k]$, where

$$\tilde{\boldsymbol{x}_\tau}[k] = \frac{B}{C} \cdot \left(\tilde{\boldsymbol{\Delta}_\tau}[k]\right)^{-1} \boldsymbol{A}_\tau \tilde{\boldsymbol{x}}[k] \tag{5.46}$$

and consequently

$$\begin{aligned} \tilde{\boldsymbol{x}_e}[k] =& \tilde{\boldsymbol{x}_c}[k] - \tilde{\boldsymbol{x}_\tau}[k] \\ =& \frac{B}{C} \Bigg\{ \left(\tilde{\boldsymbol{\Delta}_\tau}[k]\right)^{-1} \left(\tilde{\boldsymbol{\Delta}_\delta}[k]\right)^{-1} \boldsymbol{A}_\delta \tilde{\boldsymbol{x}}[k] \\ & - \left(\tilde{\boldsymbol{\Delta}_\tau}[k]\right)^{-1} \left[\mathbf{I} - \left(\tilde{\boldsymbol{\Delta}_\delta}[k]\right)^{-1}\right] \boldsymbol{A}_\tau \tilde{\boldsymbol{x}}[k] \Bigg\} \\ =& \frac{B}{C} \left(\tilde{\boldsymbol{x}_{e,1}}[k] - \tilde{\boldsymbol{x}_{e,2}}[k]\right) \end{aligned} \tag{5.47}$$

The Euclidean norm of the error term $\tilde{\boldsymbol{x}_e}[k]$ satisfies the following according to geometry inequality, writing

$$\left\|\tilde{\boldsymbol{x}_e}[k]\right\|_2 = \left\|\tilde{\boldsymbol{x}_{e,1}}[k] - \tilde{\boldsymbol{x}_{e,2}}[k]\right\|_2 \leq \left\|\tilde{\boldsymbol{x}_{e,1}}[k]\right\|_2 + \left\|\tilde{\boldsymbol{x}_{e,2}}[k]\right\|_2 \tag{5.48}$$

Examine the Euclidean norm of the first error term

$$\left\|\tilde{\boldsymbol{x}_{e,1}}[k]\right\|_2 = \left\| \left(\tilde{\boldsymbol{\Delta}_\tau}[k]\right)^{-1} \left(\tilde{\boldsymbol{\Delta}_\delta}[k]\right)^{-1} \boldsymbol{A}_\delta \tilde{\boldsymbol{x}}[k] \right\|_2 = \left\|\boldsymbol{A}_\delta \tilde{\boldsymbol{x}}[k]\right\|_2 \tag{5.49}$$

due to the fact that $\tilde{\boldsymbol{\Delta}_\tau}[k]$ and $\tilde{\boldsymbol{\Delta}_\delta}[k]$ are unitary. Recall that the assumption from Section 5.1 that the spectrum of interest has sparse channel occupancy and such occupancy pattern is hereby formally expressed by $\mathcal{K} := \operatorname{supp}\left(\tilde{\boldsymbol{x}}[k]\right)$. Moreover, without loss of generality, it is supposed that these non-zero elements in $\tilde{\boldsymbol{x}}[k]$ are independent random

variables with zero mean and variance of 1, and as a result, one can rewrite (5.49) as

$$\mathrm{E}\left[\|\boldsymbol{A}_\delta \tilde{\boldsymbol{x}}[k]\|_2^2\right] = \sum_{p=1}^{P}\sum_{q\in\mathcal{K}} \{\boldsymbol{A}_\delta\}_{p,q} \{\boldsymbol{A}_\delta\}_{p,q}^* . \tag{5.50}$$

One can safely ignore the terms in the series $\boldsymbol{A}_\delta$ whose orders are greater than 2 and simplify (5.50) as

$$\mathrm{E}\left[\|\boldsymbol{A}_\delta \tilde{\boldsymbol{x}}[k]\|_2^2\right] = 4\pi^2 \sum_{p=1}^{P}\sum_{q\in\mathcal{K}} \delta_p^2 \left(\frac{1}{2} - \frac{q-1}{C}\right)^2 , \tag{5.51}$$

on condition that δ_p's are small enough to meet

$$\left| j2\pi\delta_p \left(\frac{1}{2} - \frac{q-1}{C}\right)\frac{1}{2!}\right| \ll 1, \qquad \forall p, q, \tag{5.52}$$

which translates to

$$|\delta_p| \ll \frac{2}{\pi}, \qquad \forall p. \tag{5.53}$$

The Euclidean norm of the second term writes

$$\begin{aligned} \|\boldsymbol{x}_{\tilde{e},2}[k]\|_2 &= \left\| \left(\tilde{\boldsymbol{\Delta}}_\tau[k]\right)^{-1} \left[\mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}\right] \boldsymbol{A}_\tau \tilde{\boldsymbol{x}}[k] \right\|_2 \\ &= \left\| \left[\mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}\right] \boldsymbol{A}_\tau \tilde{\boldsymbol{x}}[k] \right\|_2 \\ &\le \left\| \mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1} \right\|_2 \|\boldsymbol{A}_\tau \tilde{\boldsymbol{x}}[k]\|_2 , \end{aligned} \tag{5.54}$$

Here, the l_2-norm for matrices refers to induced norm defined by $\|\boldsymbol{X}\|_2 := \sqrt{\lambda_{\max}\left(\boldsymbol{X}^H\boldsymbol{X}\right)}$ where $\lambda_{\max}(\cdot)$ denotes the maximum of the argument's eigenvalue, and the inequality in (5.54) is obtained by the definition of induced matrix norm.

Note that $\mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}$ is diagonal, the square of the induced-matrix-norm term $\left\| \mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1} \right\|_2$ in (5.54) is easily calculated as follows

$$\left\| \mathbf{I} - \left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1} \right\|_2^2 = \max_p \left(\left| \exp\left(j2\pi\delta_p \frac{k}{NC}\right) - 1 \right|^2 \right). \tag{5.55}$$

Similar to (5.41), one may expand $\exp\left(j2\pi\delta_p\frac{k}{NC}\right)-1$ in (5.55) by

$$\exp\left(j2\pi\delta_p\frac{k}{NC}\right)-1=\sum_{n=1}^{+\infty}\frac{\left(j2\pi\delta_p\left(\frac{k}{NC}\right)\right)^n}{n!} \tag{5.56}$$

and then only its first order term of the expansion is preserved

$$\exp\left(j2\pi\delta_p\frac{k}{NC}\right)-1=j2\pi\delta_p\left(\frac{k}{NC}\right) \tag{5.57}$$

on the condition that the skews are small enough to meet

$$\left|j2\pi\delta_p\frac{k}{NC}\frac{1}{2!}\right|\ll 1,\qquad \forall p,k, \tag{5.58}$$

which is further relaxed to the following form irrelevant to k

$$|\delta_p|\ll\frac{C}{\pi},\qquad \forall p. \tag{5.59}$$

Insert expression (5.57) back to (5.55), one finally gets the norm of the matrix norm in (5.54)

$$\left\|\mathbf{I}-\left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}\right\|_2^2=4\pi^2\left(\frac{k}{NC}\right)^2\max_p\delta_p^2. \tag{5.60}$$

The expectation of the square of the term $\|\boldsymbol{A}_\tau\tilde{\boldsymbol{x}}[k]\|_2$ in (5.54) accounts

$$\mathrm{E}\left[\|\boldsymbol{A}_\tau\tilde{\boldsymbol{x}}[k]\|_2^2\right]=\sum_{p=1}^{P}\sum_{q\in\mathcal{K}}\{\boldsymbol{A}_\tau\}_{p,q}\{\boldsymbol{A}_\tau\}_{p,q}^*=|\mathcal{K}|\cdot P. \tag{5.61}$$

Consequently, according to (5.60) and (5.61), the expectation of (5.54) writes

$$\begin{aligned}\mathrm{E}\left[\|\tilde{\boldsymbol{x}_{e,2}}[k]\|_2^2\right]&\le\left\|\mathbf{I}-\left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}\right\|_2^2\cdot\mathrm{E}\left[\|\boldsymbol{A}_\tau\tilde{\boldsymbol{x}}[k]\|_2^2\right]\\&=|\mathcal{K}|\,P\cdot\left\|\mathbf{I}-\left(\tilde{\boldsymbol{\Delta}}_\delta[k]\right)^{-1}\right\|_2^2\\&=4\pi^2\,|\mathcal{K}|\,P\left(\frac{k}{NC}\right)^2\cdot\max_p\delta_p^2\end{aligned} \tag{5.62}$$

where the equality in (5.54) and (5.62) is met when the diagonal terms in $\mathbf{I} - \left(\tilde{\mathbf{\Delta}}_{\delta}[k]\right)^{-1}$ are equal, i.e. $\delta_1 = \delta_2 = \cdots = \delta_P$.

According to Cauchy-Schwarz inequality in the form of expectations of random variable, one has the following property on the expectation of arbitrary random variables $X, Y \in \mathbb{C}$,

$$\left|\mathrm{E}\left[XY\right]\right|^2 \leq \mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right], \tag{5.63}$$

which immediately leads to the following

$$\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e,1}}[k]\right\|_2 \left\|\tilde{\boldsymbol{x}_{e,2}}[k]\right\|_2\right] \leq \sqrt{\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e,1}}[k]\right\|_2^2\right]\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e,2}}[k]\right\|_2^2\right]}. \tag{5.64}$$

From (5.47), finally one has the upper bound of the expectation of error vector by scaling using inequalities (5.48), (5.62) and (5.64),

$$\begin{aligned}
\sigma^2_{\tilde{\boldsymbol{x}}_e[k]} &:= \mathrm{E}\left[\left\|\tilde{\boldsymbol{x}}_e[k]\right\|_2^2\right] = \frac{B^2}{C^2}\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e1}}[k] + \tilde{\boldsymbol{x}_{e2}}[k]\right\|_2^2\right] \\
&\leq \frac{B^2}{C^2}\mathrm{E}\left[\left(\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2 + \left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2\right)^2\right] \\
&= \frac{B^2}{C^2}\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2^2 + \left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2^2 + 2\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2\left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2\right] \\
&\leq \frac{B^2}{C^2}\Bigg(\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2^2\right] + \mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2^2\right] \\
&\qquad + 2\sqrt{\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2^2\right]\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2^2\right]}\Bigg) \\
&= \frac{B^2}{C^2}\left(\sqrt{\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e1}}[k]\right\|_2^2\right]} + \sqrt{\mathrm{E}\left[\left\|\tilde{\boldsymbol{x}_{e2}}[k]\right\|_2^2\right]}\right)^2 \\
&\leq \left(\frac{2\pi B}{C}\right)^2\Bigg(\sqrt{|\mathcal{K}|\,P}\frac{k}{NC}\cdot\max_p|\delta_p| \\
&\qquad + \sqrt{\sum_{p=1}^{p}\delta_p^2\sum_{q\in\mathcal{K}}\left(\frac{1}{2} - \frac{q-1}{C}\right)^2}\Bigg)^2.
\end{aligned} \tag{5.65}$$

The power of the signal without the impact of timing skew is denoted by

$$\sigma^2_{\tilde{\boldsymbol{x}}_\tau[k]} := \mathrm{E}\left[\left\|\tilde{\boldsymbol{x}}_\tau[k]\right\|_2^2\right] = \frac{B^2}{C^2}|\mathcal{K}|\,P \tag{5.66}$$

as a direct result of (5.61). Define the error-to-signal ratio (ESR) of the measured signal $\tilde{\boldsymbol{x}_\tau}[k]$ by

$$ESR_{\tilde{\boldsymbol{x}_c}[k]} := \frac{\sigma^2_{\tilde{\boldsymbol{x}_e}[k]}}{\sigma^2_{\tilde{\boldsymbol{x}_\tau}[k]}}, \tag{5.67}$$

which is considered as a metric describing the level of error appearing in the measurements compared to the signal power. The upper bound of ESR is obtained from (5.65)

$$ESR_{\tilde{\boldsymbol{x}_c}[k]} \leq 4\pi^2 \left(k\rho_1\psi_1 + \frac{\rho_2\psi_2}{\sqrt{|\mathcal{K}|\, P}} \right)^2 \tag{5.68}$$

where for clarity, denote $\rho_1 := \frac{1}{NC}$ and the term related to channel occupancy is presented by $\rho_2 := \sqrt{\sum_{q\in\mathcal{K}} \left(\frac{1}{2} - \frac{q-1}{C}\right)^2}$, and ρ_1 and ρ_2 are separated in two terms. The skew-related terms are expressed by $\psi_1 := \max_p |\delta_p|$ and $\psi_2 := \sqrt{\left(\sum_{p=1}^{p} \delta_p^2\right)}$.

5.5.1.2 Error in MMV Measurements

With Unknown Skews If the skews of cosets are unknown, having obtained the DFT of each coset's samples $\tilde{\boldsymbol{x}_c}[k]$'s, one can only construct the MMV measurements $\tilde{\boldsymbol{y}}[k]$'s in (5.12) using the known τ_p-related $\tilde{\boldsymbol{\Delta}}_\tau[k]$ in (5.39), writing

$$\tilde{\boldsymbol{y}}[k] = \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}_\tau[k]\tilde{\boldsymbol{x}_c}[k] = \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}_\tau[k] \left(\tilde{\boldsymbol{x}_\tau}[k] + \tilde{\boldsymbol{x}_e}[k]\right). \tag{5.69}$$

Denote the skew-free version of MMV measurement $\tilde{\boldsymbol{y}}[k]$ by

$$\tilde{\boldsymbol{y}_\tau}[k] := \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}_\tau[k]\tilde{\boldsymbol{x}_\tau}[k], \tag{5.70}$$

and the error in MMV measurement is denoted by

$$\tilde{\boldsymbol{y}_e}[k] := \tilde{\boldsymbol{y}}[k] - \tilde{\boldsymbol{y}_\tau}[k] = \frac{C}{B}\tilde{\boldsymbol{\Delta}}_\tau[k]\tilde{\boldsymbol{x}_e}[k], \tag{5.71}$$

which can be alternatively expressed in the matrix form by stacking these N column vectors

$$\boldsymbol{Y}_e = \boldsymbol{Y} - \boldsymbol{Y}_\tau. \tag{5.72}$$

Define the power of error signal $\tilde{\boldsymbol{y}_e}[k]$, expressed by

$$\sigma^2_{\tilde{\boldsymbol{y}_e}[k]} := \mathrm{E}\left[\|\tilde{\boldsymbol{y}_e}[k]\|_2^2\right] = \frac{C^2}{B^2}\sigma^2_{\tilde{\boldsymbol{x}_e}[k]}. \tag{5.73}$$

The power of skew-free signal $\tilde{\boldsymbol{y}_\tau}[k]$ writes

$$\sigma^2_{\tilde{\boldsymbol{y}_\tau}[k]} := \mathrm{E}\left[\|\tilde{\boldsymbol{y}_\tau}[k]\|_2^2\right] = \frac{C^2}{B^2}\mathrm{E}\left[\|\tilde{\boldsymbol{x}_\tau}[k]\|_2^2\right] = \frac{C^2}{B^2}\sigma^2_{\tilde{\boldsymbol{x}_\tau}[k]}. \tag{5.74}$$

For each measurement vector in $\boldsymbol{Y}$, i.e. $\tilde{\boldsymbol{y}}[k]$ for $k = 0, 1, \cdots, N-1$, define the ESR as follows

$$ESR_{\tilde{\boldsymbol{y}}[k]} := \frac{\sigma^2_{\tilde{\boldsymbol{y}_e}[k]}}{\sigma^2_{\tilde{\boldsymbol{y}_\tau}[k]}} = \frac{\sigma^2_{\tilde{\boldsymbol{x}_e}[k]}}{\sigma^2_{\tilde{\boldsymbol{x}_\tau}[k]}} = ESR_{\tilde{\boldsymbol{x}}[k]} \tag{5.75}$$

Note Frobenius-norm of matrix by $\|\boldsymbol{X}\|_F$ and by definition the power of the matrix-form signal (5.72) is given by $\sigma^2_{\boldsymbol{Y}_e} := \mathrm{E}\left[\|\boldsymbol{Y}_e\|_F^2\right]$ and $\sigma^2_{\boldsymbol{Y}_\tau} := \mathrm{E}\left[\|\boldsymbol{Y}_\tau\|_F^2\right]$. For the matrix $\boldsymbol{Y}$ of measurements, define the ESR as follows

$$\begin{aligned} ESR_{\boldsymbol{Y}} &:= \frac{\sigma^2_{\boldsymbol{Y}}}{\sigma^2_{\boldsymbol{X}}} = \frac{\sum_{k=0}^{N-1}\sigma^2_{\tilde{\boldsymbol{y}_e}[k]}}{\sum_{k=0}^{N-1}\sigma^2_{\tilde{\boldsymbol{y}_\tau}[k]}} = \frac{1}{N}\sum_{k=0}^{N-1} ESR_{\tilde{\boldsymbol{y}}[k]} \\ &\le 4\pi^2 \cdot \left(\frac{4\psi_1^2}{3C^2}\frac{(N-1)(N-\frac{1}{2})}{N^2} \right. \\ &\qquad \left. + \frac{\rho_2^2\psi_2^2}{|\mathcal{K}|\,P} + \frac{\rho_2\psi_1\psi_2}{C\sqrt{|\mathcal{K}|\,P}}\frac{(N-1)(N-2)}{N^2} \right) \\ &< 4\pi^2 \cdot \left(\frac{2\psi_1^2}{3C^2} + \frac{\rho_2^2\psi_2^2}{|\mathcal{K}|\,P} + \frac{\rho_2\psi_1\psi_2}{C\sqrt{|\mathcal{K}|\,P}} \right). \end{aligned} \tag{5.76}$$

With Known Skews If the skews are known, the MMV measurements are obtained as in (5.12) using $\tilde{\boldsymbol{\Delta}}[k]$ which relates to both known offsets τ_p's and skews δ_t's. Recall

(5.12) and separate $\tilde{\boldsymbol{\Delta}}[k]$ according to (5.39)

$$\begin{aligned}\tilde{\boldsymbol{y}}[k] &= \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}[k]\tilde{\boldsymbol{x}_c}[k] \\ &= \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}}_\delta[k]\tilde{\boldsymbol{\Delta}}_\tau[k]\tilde{\boldsymbol{x}_c}[k] = (\boldsymbol{A}_\tau + \boldsymbol{A}_\delta)\,\tilde{\boldsymbol{x}}[k].\end{aligned} \tag{5.77}$$

By relating (5.46) in the definition the skew-free version of MMV measurement $\tilde{\boldsymbol{y}}[k]$, it is then expressed by

$$\tilde{\boldsymbol{y}_\tau}[k] := \frac{C}{B} \cdot \tilde{\boldsymbol{\Delta}_\tau}[k]\tilde{\boldsymbol{x}_\tau}[k] = \boldsymbol{A}_\tau\tilde{\boldsymbol{x}}[k]. \tag{5.78}$$

Consequently, the error in MMV measurement writes

$$\tilde{\boldsymbol{y}_e}[k] = \tilde{\boldsymbol{y}}[k] - \tilde{\boldsymbol{y}_\tau}[k] = \boldsymbol{A}_\delta\tilde{\boldsymbol{x}}[k]. \tag{5.79}$$

The power of error signal $\tilde{\boldsymbol{y}_e}[k]$ in the scenario of known skews writes as follows according to (5.50)

$$\sigma^2_{\tilde{\boldsymbol{y}_e}[k]} := \mathrm{E}\left[\|\tilde{\boldsymbol{y}_e}[k]\|_2^2\right] = 4\pi^2\psi_2^2\rho_2^2. \tag{5.80}$$

The power of skew-free signal $\tilde{\boldsymbol{y}_\tau}[k]$ in the scenario of known skews writes as follows according to (5.61)

$$\sigma^2_{\tilde{\boldsymbol{y}_\tau}[k]} := \mathrm{E}\left[\|\tilde{\boldsymbol{y}_\tau}[k]\|_2^2\right] = |\mathcal{K}|\,P. \tag{5.81}$$

For each measurement vector in $\boldsymbol{Y}$, i.e. $\tilde{\boldsymbol{y}}[k]$ for $k = 0, 1, \cdots, N-1$, the ESR of $\tilde{\boldsymbol{y}}[k]$ is hereby expressed by

$$ESR_{\tilde{\boldsymbol{y}}[k]} := \frac{\sigma^2_{\tilde{\boldsymbol{y}_e}[k]}}{\sigma^2_{\tilde{\boldsymbol{y}_\tau}[k]}} = 4\pi^2 \cdot \frac{\psi_2^2\rho_2^2}{|\mathcal{K}|\,P}. \tag{5.82}$$

Furthermore, for the matrix $\boldsymbol{Y}$ of measurements, the ESR of $\boldsymbol{Y}$ is expressed by

$$ESR_{\boldsymbol{Y}} := \frac{\sum_{k=0}^{N-1}\sigma^2_{\tilde{\boldsymbol{y}}_e[k]}}{\sum_{k=0}^{N-1}\sigma^2_{\tilde{\boldsymbol{y}}_\tau[k]}} = 4\pi^2 \cdot \frac{\psi_2^2\rho_2^2}{|\mathcal{K}|\,P} \tag{5.83}$$

which is an exact value instead of an upper bound.

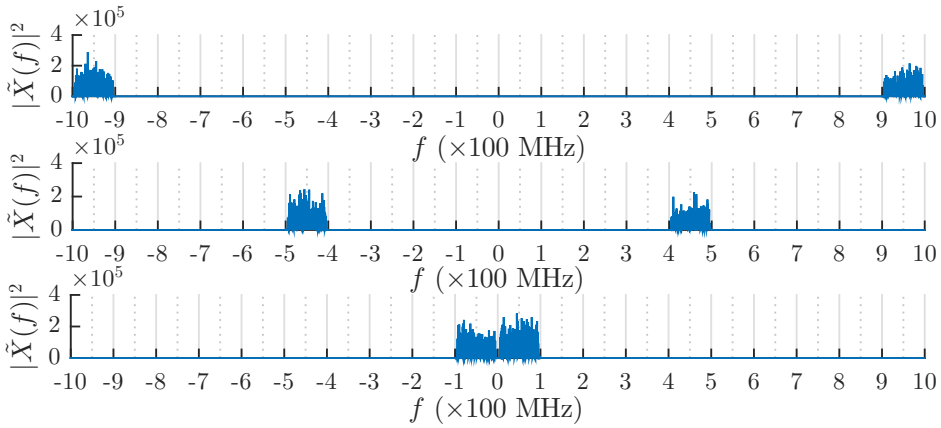


Figure 5.8: Frequency-domain representation of generated baseband waveforms in Section 5.5.2, for channel occupancy pattern I, II and III from top down.

5.5.2 Numeric Analysis

Real-world baseband signals generated by a National Instruments mmWave SDR platform are used for evaluation. The signals span the frequency range of -1 GHz to 1 GHz, i.e. $B = 2$ GHz, and frames of $I = 81920$ raw samples from an ADC sampling at 3.072 GHz are collected, denoted by $r[i]$'s. To obtain the continuous-time signal $s(t)$, one may perform interpolation using a sinc function to preserve the components within the 2 GHz instantaneous bandwidth by

$$s(t) = \sum_{i=0}^{I} 2B \cdot r[i] \mathrm{sinc}\left(2Bt\right).$$

Split the baseband bandwidth into $C = 40$ channels so that each channel spans $B/C = 50$ MHz, and the SDR transmitter can generate multiple 100 MHz orthogonal frequency-division multiplexing waveforms at given central frequencies. Signal frames of three-channel occupancy patterns are used, referred to as pattern I, II, and III, as shown in Fig. 5.8. To remove the undesired impact of additive thermal noise in this analysis, noise components in the vacant channels are zeroed in these frames. In all three patterns, four channels are occupied, i.e. $|\mathcal{K}| = 4$, and channel occupancy is $\mathcal{K} = \{1, 2, 39, 40\}$, $\{11, 12, 29, 30\}$, and $\{19, 20, 21, 22\}$ for pattern I, II, and III respectively. The number of samples per coset for each frame is set to $N = 4096$.

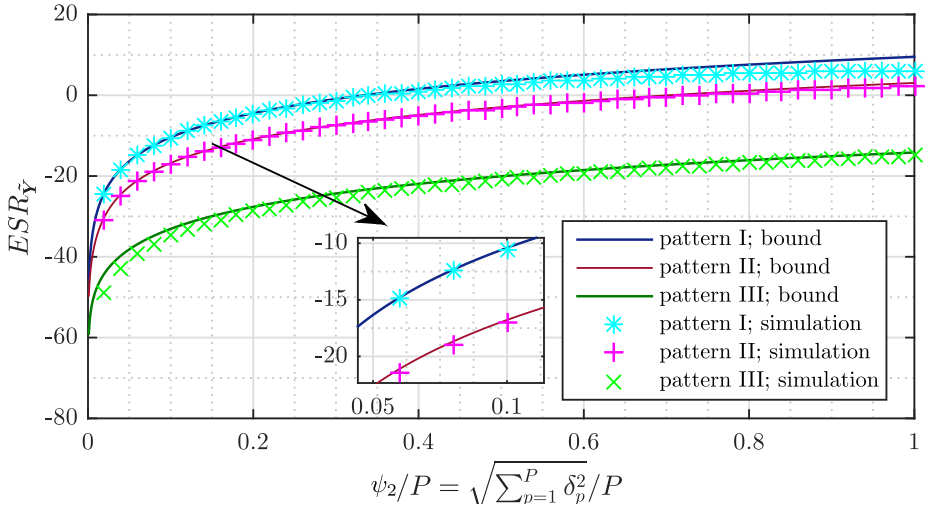


Figure 5.9: ESR bounds in (5.76) and emprical ESRs from signals of three channel occupancy patterns against the level of unknown skews.

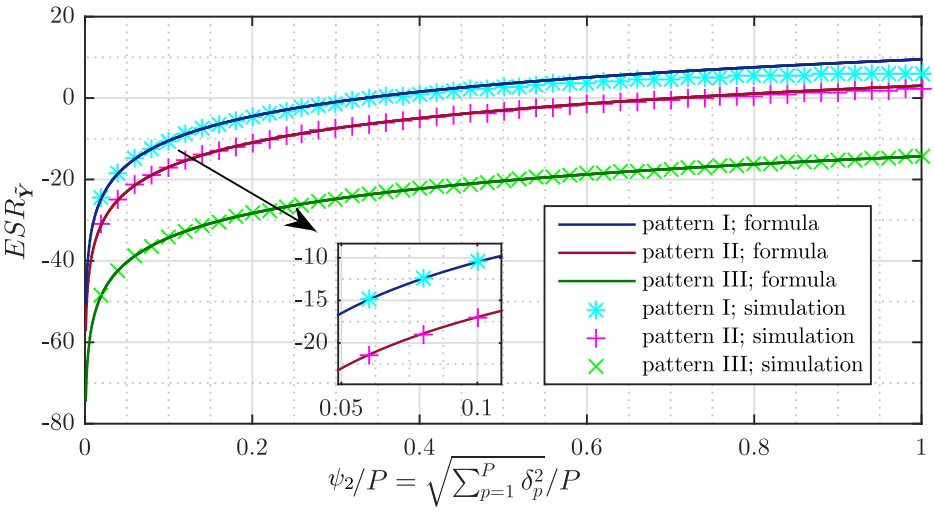


Figure 5.10: Analytic-form ESR in (5.83) and emprical ESRs from signals of three channel occupancy patterns against the level of known skews.

In Fig. 5.9, under the circumstance of unknown skews, the ESR bounds of three occupancy patterns calculated by (5.76) are plotted against the level of skews across all cosets, quantified by the mean of root sum square of skews, i.e. $\psi_2/P = \sqrt{\sum_{p=1}^{P} \delta_p^2}/P$. Empirical ESRs are calculated from 500 sample frames in the SDR-generated dataset using various levels of skews, which are also shown in Fig. 5.9 as scatters. It can be seen that each curve from (5.76) serves as the upper bound of the empirical ESRs correctly. Moreover, it is also noted that the ESR significantly varies with the channel

occupancy pattern. Specifically, the more distant the occupied channels are from the central frequency, the larger ESR it leads to, with the same amount of skew, which can be ascribed to the occupancy-related term ρ_2 in (5.76).

In Fig. 5.10, for known skews, the analytic-form ESR and also the empirical ESRs of three occupancy patterns calculated by (5.83) against the level of skews across all cosets are presented, ψ_2/P. It is found that the analytic-form ESR curves align well with the simulation results when skews are small. Furthermore, when skews are large, the high-order terms in the Taylor expansion (5.41) becomes too evident to be ignored, and consequently, the analytical ESR curves deviate from the simulation results. One may include higher-order terms in (5.41) to have a more accurate but complex form of the analytical ESR.

Finally, the recovery performance of SDR-generated signals of three occupancy patterns is illustrated in Fig. 5.11. The simultaneous orthogonal matching pursuit (SOMP) [174] is implemented as the recovery algorithm for the MMV model (5.13) and examine the normalized mean square error (NMSE) and detection probability of occupied channels in Fig. 5.11 (a) and (b) respectively, where NMSE is formally defined by $NMSE := ||\tilde{\boldsymbol{X}}_r - \tilde{\boldsymbol{X}}||_F^2/||\tilde{\boldsymbol{X}}||_F^2$, where $\tilde{\boldsymbol{X}}_r$ denotes the signal recovered by SOMP. Note that both NMSE and detection probability gets deteriorated with increasing level of skews, and with channel occupancy more distant from the central frequency. The trends of skew-induced recovery performance deterioration are consistent with increasing ESRs.

5.6 Summary

This chapter focuses on the multicoset sampler architecture as the front-end of a CSS system in wideband scenarios for mmWave applications and provides the modeling of practical multicoset sampler as an MMV model. In the following, two major problems emerging in the attempts of applying low-complexity greedy algorithms to the multi-

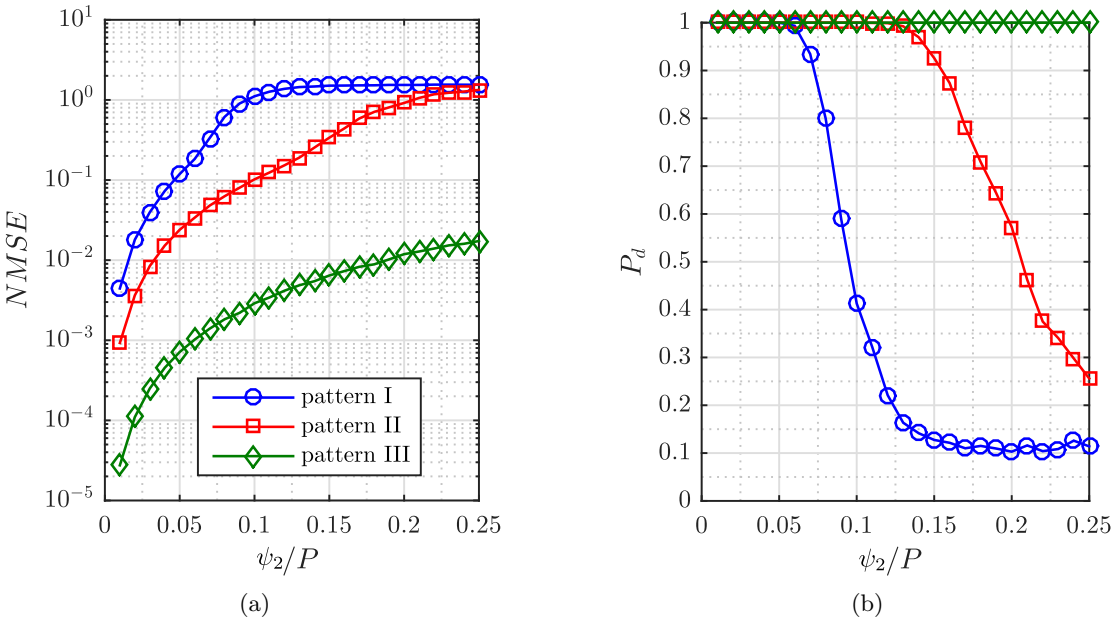


Figure 5.11: Recovery performance of signals of three-channel occupancy patterns using SOMP against the level of skews. **(a)** NMSE; **(b)** Detection probability.

coset sampler model are discussed. Firstly, the spectrum leakage problem is considered and a windowing scheme is proposed for the multicoset sampler model to diminish the spectrum leakage compact in the spectrum sparsity estimation and the sparse recovery of the spectrum. Then, to obtain the necessary input for the greedy algorithms, a spectrum sparsity estimation scheme is proposed. Moreover, with such sparsity information, the original MMV CS model can be simplified by subspace techniques, and hence the sparse recovery complexity can be reduced. Next, an SDR demo for a mmWave-band multicoset-sampler-based CSS system is presented. Serving as a testbed for a realistic multi-gigahertz bandwidth CSS system, the demo is developed on the NI mmWave platform and achieves multicoset sampling with sparse recovery by greedy algorithms and the abovementioned windowing and sparsity estimation schemes.

Furthermore, it is noted that one major technical challenge for real-world wideband multicoset sampler is the timing control of the sampling, especially when the bandwidth reaches multiple gigahertz in mmWave applications. Inaccuracies in sampling timing

would induce undesired errors in the samples taken by the multicoset sampler. Since such errors further contribute to the inaccuracies in the recovered spectrum for sparse recovery techniques, it is of interest at the system level to quantify such measurement errors. In Section 5.5 the timing-skew-induced error in the MMV measurements of the multicoset sampler is analyzed.

Chapter 6

Conclusion and Future Work

6.1 Conclusions

The ever-evolving development for wireless communication services and the increasing demand for high data rates and wide service coverage has been facing the bottleneck of spectrum scarcity. The availability of the finite spectrum resource with certain propagation capabilities has been a persisting need. Spectrum sensing and cognitive radio techniques are essential to obtain accurate information of ambient radio activities and spectrum usage, enabling the vacant spectrum to be used in dynamic spectrum access applications. With the recent developments of compressive sensing theories, sub-Nyquist techniques have been applied for the application of real-time wideband spectrum sensing to massively reduce the necessary sampling rate, on which topics this thesis focuses. The background and current developments of compressive sensing techniques in the context of wideband spectrum sensing are reviewed in Chapter 2. In the following chapters, some major work during the Ph.D. has been presented contributing to addressing the following three challenges in the current development of compressive spectrum sensing -

- the low-complexity realization of recovery of the sparse spectrum and the sparsity estimation issue;

- channel energy modeling and robust energy detection for occupied channels in the wideband spectrum;

- realizing the multicoset sampler architecture as the sub-Nyquist sampling front end towards gigahertz bandwidth for mmWave applications and identifying the imperfections due to technical limitations.

In Chapter 3, the low-computational-complexity recovery of spectrum in compressive spectrum sensing is focused on. It is proposed that the key problem of sparsity estimation for low-complexity greedy algorithms can be solved directly and rapidly by an eigendecomposition-based detector. By further formulating a reduced-dimension compressive sensing model to be solved by a chosen greedy algorithm, it is presented that computational complexity can be greatly saved and the performance of incumbent radio detection enhanced as a result of the aid of sparsity estimation and dimension reduction.

In Chapter 4, it is discovered that the statistical model of the recovered power spectrum by CSS is different from that of the original signal for conventional spectrum sensing. This part of the work especially considers the energy detection problem and tentatively builds the channel energy statistical model for CSS. Additionally, based on the statistical model, a threshold adaption scheme is proposed for CSS to keep a constant false alarm rate in the energy detection.

In Chapter 5, the multicoset sampler architecture is examined in particular towards its real-world implementation. A few practical problems of the processing routines are discussed, such as the windowing of the measurements and low-complexity sparse recovery. Next, a demonstration for a wideband CSS system emulating multicoset sampler at 28.5 GHz is built on the National Instruments mmWave soft-defined-radio platform. Furthermore, it is identified that the timing accuracy of the sampling is essential for multicoset sampler in wideband applications and the initial investigation is made to quantify the timing-skew-induced error in the compressed measurements.

6.2 Future Work

Although the theory of compressive sensing and various signal processing topics in wideband CSS have been intensively investigated, open problems remain to be addressed towards the real-world implementations of spectrum sensing with wider bandwidth, considering the distinctive architectures of sub-Nyquist samplers and the practical limitations. The promising topics of future work are identified as follows.

6.2.1 On the Hardware Mismatches of Sub-Nyquist Samplers

Multicoset sampler consists of an ADC array to achieve interleaved sampling patterns. However, one major problem with time interleaving sampling is that the ADCs and the parallel analog signal paths before the ADCs, though required to be identical, do not practically have identical performances. Such part-to-part discrepancies refer to the properties such as gains, phase offsets, dispersion, and in-phase-quadrature imbalance. For conventional Nyquist-rate interleaved sampling systems, these mismatches cause non-harmonic spurious impairments on the frequency domain of the sampled signal, known as 'interleaving spurs', which has been frequently studied and considered in practice [204]. However, the multicoset sampler inheriting the interleaved sampling structure and also the modulated wideband converter (MWC) with parallel ADCs would suffer from such mismatches as well, and it is hence necessary to model the mismatches for the multicoset sampler, quantify the error in the measurements or the recovered signal. Moreover, existing conventional foreground [205] (using a known calibration signal) and background calibration techniques [206, 207] have seen to apply to Nyquist-rate interleaved ADCs with little extra hardware, and similar strategies can potentially adapt to the sub-Nyquist sampling models to overcome such hardware imparities.

6.2.2 Compensating the Undesired Skews of Multicoset Sampler

It is indicated in Chapter 5 that the undesired timing skews can cause a significant impact on the accuracy of the compressed measurements. Although some of these undesired

skews can be minimized on the hardware, i.e. precise length matching of the traces and analog delay units, also there may be possibilities to compensate these skew on the signal processing level. The current multicoset sampler model dictates the timing offsets to have the resolution of the Nyquist sampling period so that the sensing matrix is a partial Fourier matrix where the rows are ideally orthogonal. It is intriguing to know the system performance if one includes the undesired skews into the sensing matrix to form non-ideally-orthogonal rows and compare with the performance with unknown skews and the impaired measurements. By inserting the undesired skew terms in the compressive sensing model, such an idea of compensation may reduce the impact of these undesired skews.

6.2.3 Quantifying the Skew-Induced Error in the Recovered Signal and Other Sub-Nyquist Samplers

The skew-induced error on the compressed measurement of the multicoset sampler is quantified in Section 5.5. However, for a more direct indicator of system-level performance deterioration, one may account for the skew-induced error in the final signal recovered by a specific sparse recovery algorithm by analyzing the sparse recovery routines, potentially via some underlying properties such as the restricted isometry properties. It is also noted that the undesired skew is not only a problem specifically for the multicoset sampler, a similar issue arises for other sub-Nyquist sampling front-ends employing parallel branches and requiring synchronization, such as the MWC. It is hence worth investigating the skew-induced error issue for other sub-Nyquist samplers for the purpose to implement them for a larger bandwidth.

6.2.4 Performances of Windowing Schemes for Sub-Nyquist Samplers

It is proposed in Section 5.2 that a Dolph-Chebyshev window can be used in particular to mitigate the impact of the spectrum leakage to the sparsity of the signal and hence the eigendecomposition-based sparsity detector in the context of the multicoset sampler. However, when dealing with finite-length data, other types of common windowing func-

tions may still be interested to be applied for varying levels of compromise between the windowed signal's frequency resolution and amplitude resolution (or dynamic range), and their empirical performances of sparse recovery and implications to the sparsity estimation may need to be evaluated. More generally, the use of various types of windows with steep roll-off frequency responses for sub-Nyquist sampling is especially worth investigating, as general compressive sensing theories do not necessitate a strictly sparse spectrum model for robust sparse recovery, that is, fast-decaying non-ideally-sparse signals can be reconstructed into a sparse approximate with small reconstruction error [208]. Examples can be seen in [73] that in the MWC setting, a violation of the strict sparsity model caused by spectrum leakage, even with rectangular windowing, does not necessarily fail the signal reconstruction.

6.2.5 Adaptive Adjustment of Compression Ratio in Compressive Spectrum Sensing

The compression of the sampling rate in CSS is desired to be as low as possible to minimize the power consumption while ensuring the sparse recovery performance. In Section 2.4.3 it is mentioned that it is sufficient to use a sensing matrix to recover an arbitrary sparse signal of which the sparsity is less than half of the Kruskal rank of the sensing matrix. In the context of CSS, for the multicoset sampler and the MWC model, this condition translates to the following - it is sufficient to have the number of parallel ADC branches larger than two times the number of the occupied channels of the spectrum to ensure exact sparse recovery. One may devise an adaption scheme to determine the number of active ADCs (i.e. the compression ratio) sufficient for reliable spectrum recovery based on the sparsity estimation technique proposed in this thesis. Additionally, it may also be possible to use some comparison metrics between the raw sub-Nyquist samples from the ADCs and the recovered version to evaluates the recovery performance and tentatively adjust the number of active ADCs based on such empirical metrics.

References

[1] Y. Gao, Z. Qin, Z. Feng, Q. Zhang, O. Holland, and M. Dohler, "Scalable and reliable IoT enabled by dynamic spectrum management for M2M in LTE-A," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1135–1145, 2016.

[2] K. Chatzikokolakis, P. Spapis, A. Kaloxylos, and N. Alonistioti, "Toward spectrum sharing: opportunities and technical enablers," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 26–33, 2015.

[3] R. N. Clarke, "Expanding mobile wireless capacity: The challenges presented by technology and economics," *Telecommunications Policy*, vol. 38, no. 8, pp. 693 – 708, 2014.

[4] O. Holland, S. Ping, A. Aijaz, J. M. Chareau, P. Chawdhry, Y. Gao, Z. Qin, and H. Kokkinen, "To white space or not to white space: That is the trial within the OfCom TV white spaces pilot," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Sept. 2015, pp. 11–22.

[5] "Implementing TV white spaces," Office of Communications, U.K., Tech. Rep., Dec. 2015. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0034/68668/tvws-statement.pdf [Accessed: 2020-05-31]

[6] "Statement on licence-exempting cognitive devices using interleaved spectrum," National Instruments, Tech. Rep., Jul. 2009. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/40838/statement.pdf [Accessed: 2020-05-31]

[7] Y. Ma, Y. Gao, C. Fu, W. Rong, Z. Xiong, and S. Cui, "TV white space spectrum analysis based on machine learning," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 68–80, Jun. 2019.

[8] "Report and order on consideration," Federal Communications Commision, Tech. Rep., Mar. 2019. [Online]. Available: https://docs.fcc.gov/public/attachments/FCC-19-24A1.pdf [Accessed: 2020-05-31]

[9] "3.5 GHz band overview," Federal Communications Commission, Tech. Rep., Feb. 2020. [Online]. Available: https://www.fcc.gov/wireless/bureau-divisions/

mobility-division/35-ghz-band/35-ghz-band-overview [Accessed: 2020-05-31]

[10] A. Attar, O. Holland, and H. Aghvami, "Chapter 5 - spectrum access and sharing," in *Cognitive Radio Communications and Networks*, A. M. Wyglinski, M. Nekovee, and Y. T. Hou, Eds. Oxford: Academic Press, 2010, pp. 113 – 148.

[11] J. Wang, M. Ghosh, and K. Challapali, "Emerging cognitive radio applications: A survey," *IEEE Communications Magazine*, vol. 49, no. 3, pp. 74–81, 2011.

[12] M. Abo-Zahhad, S. Ahmed, M. Asaduzzaman Farrag, and K. Ali, "Wideband cognitive radio networks based compressed spectrum sensing: A survey," *Journal of Signal and Information Processing*, vol. 09, pp. 122–151, Jan. 2018.

[13] J. J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, "Collaborative spectrum sensing from sparse observations in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 327–337, Feb. 2011.

[14] Z. Qin, Y. Gao, and C. G. Parini, "Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1174–1185, Feb. 2016.

[15] Z. Qin, Y. Gao, M. D. Plumbley, and C. G. Parini, "Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3106–3117, Jun. 2016.

[16] Y. Ma, Y. Gao, Y. C. Liang, and S. Cui, "Reliable and efficient sub-Nyquist wideband spectrum sensing in cooperative cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2750–2762, Oct. 2016.

[17] "TS 38.101-1: NR; user equipment (UE) radio transmission and reception; part 1: Range 1 standalone (Ver. 16.3.0)," 3GPP, Tech. Rep., 2020. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.101-1/ [Accessed: 2020-06-29]

[18] "TS 38.101-2: NR; user equipment (UE) radio transmission and reception; part 2: Range 2 standalone (Ver. 16.3.1)," 3GPP, Tech. Rep., 2020. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.101-2/ [Accessed: 2020-06-29]

[19] "Our decision to implement technical and regulatory changes to the 57 – 71 GHz band," Office of Communications (UK), Tech. Rep. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0013/126121/Statement_Implementing-Ofcoms-decision-on-the-57-71GHz-band.pdf [Accessed: 2020-05-31]

[20] "Fact sheet: Spectrum frontiers proposal to identify, open up vast amounts of new high-band spectrum for next generation (5G) wireless broadband," Federal Communications Commission (US), Tech. Rep. [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-340310A1.pdf [Accessed: 2020-05-31]

[21] M. Rebato and M. Zorzi, "A spectrum sharing solution for the efficient use of mmwave bands in 5G cellular scenarios," in *IEEE International Symposium Dynamic Spectrum Access Networks (DySPAN)*, Aug. 2018.

[22] J. Mitola, "Cognitive radio - an integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Institute of Technology, Kista, Sweden, 2000.

[23] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications Magazine*, vol. 6, no. 4, pp. 13–18, 1999.

[24] Y. Liang, K. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: an overview," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3386–3407, 2011.

[25] I. F. Akyildiz, W. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, 2008.

[26] M. Fitch, M. Nekovee, S. Kawade, K. Briggs, and R. MacKenzie, "Wireless service provision in TV white space with cognitive radio technology: A telecom operator's perspective and experience," *IEEE Communications Magazine*, vol. 49, no. 3, pp. 64–73, 2011.

[27] C. Cordeiro, K. Challapali, D. Birru, and Sai Shankar, "IEEE 802.22: the first

worldwide wireless standard based on cognitive radios," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2005, pp. 328–337.

[28] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, "IEEE 802.11af: a standard for TV white space spectrum sharing," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, 2013.

[29] M. M. Sohul, M. Yao, T. Yang, and J. H. Reed, "Spectrum access system for the citizen broadband radio service," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 18–25, 2015.

[30] B. Ren, M. Wang, J. Zhang, W. Yang, J. Zou, M. Hua, and X. You, "Cellular communications on license-exempt spectrum," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 146–153, 2016.

[31] M. Labib, V. Marojevic, J. H. Reed, and A. I. Zaghloul, "Extending LTE into the unlicensed spectrum: Technical analysis of the proposed variants," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 31–39, 2017.

[32] "LTE and Wi-Fi in unlicensed spectrum: A coexistence study," Google Inc., Tech. Rep., 2015. [Online]. Available: https://drive.google.com/file/d/0B-iq0rxkyMLyVUxkNWtoUzItZDA/view [Accessed: 2020-05-31]

[33] "Study on NR-based access to unlicensed spectrum (Ver. 16.0.0)," 3GPP, Tech. Rep., 2018. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.889/ [Accessed: 2020-06-29]

[34] "3GPP TSG-RAN 86 meeting, extending current NR operation to 71 GHz," 3GPP, Tech. Rep., 2019. [Online]. Available: https://www.3gpp.org/ftp/Meetings_3GPP_SYNC/RAN/Inbox_history/4_Thursday_noon/RP-193229.zip [Accessed: 2020-06-29]

[35] N. Patriciello, S. Lagén, B. Bojović, and L. Giupponi, "NR-U and IEEE 802.11 technologies coexistence in unlicensed mmWave spectrum: Models and evaluation," *IEEE Access*, vol. 8, pp. 71 254–71 271, 2020.

[36] S. Lagen, L. Giupponi, S. Goyal, N. Patriciello, B. Bojović, A. Demir, and M. Beluri, "New radio beam-based access to unlicensed spectrum: Design challenges and solu-

tions," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 1, p. 8–37, Jan. 2020.

[37] "Use of spectrum bands above 24 GHz for mobile radio services," Federal Communications Commission, Tech. Rep., Oct. 2017. [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-347449A1.pdf [Accessed: 2020-05-31]

[38] "Fact sheet: Spectrum frontiers rules identify, open up vast amounts of new high-band spectrum for next generation (5G) wireless broadband," Federal Communications Commission, Tech. Rep., Jun. 2016. [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-339990A1.pdf [Accessed: 2020-05-31]

[39] F. Guidolin, M. Nekovee, L. Badia, and M. Zorzi, "A study on the coexistence of fixed satellite service and cellular networks in a mmWave scenario," in *IEEE International Conference on Communications (ICC)*, 2015.

[40] F. Guidolin and M. Nekovee, "Investigating spectrum sharing between 5G millimeter wave networks and fixed satellite systems," in *IEEE Global Communications Conference (GLOBECOM) Workshops*, 2015.

[41] A. Ghasemi and E. S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Network (DySPAN)*, Nov. 2005, pp. 131–136.

[42] W. Han, J. Li, Z. Tian, and Y. Zhang, "Efficient cooperative spectrum sensing with minimum overhead in cognitive radio," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3006–3011, Oct. 2010.

[43] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[44] O. Holland, N. Sastry, S. Ping, R. Knopp, F. Kaltenberger, D. Nussbaum, J. Hallio, M. Jakobsson, J. Auranen, R. Ekman, J. Paavola, A. Kivinen, H. Tran, K. Ishizu, H. Harada, P. Chawdhry, J. Chareau, J. Bishop, M. Bavaro, E. Anguili, Y. Gao, R. Dionisio, P. Marques, H. Kokkinen, and O. Luukkonen, "A series of trials in

the UK as part of the OfCom TV white spaces pilot," in *International Workshop on Cognitive Cellular Systems (CCS)*, 2014, pp. 1–5.

[45] Y. Ma, X. Zhang, and Y. Gao, "Joint sub-Nyquist spectrum sensing scheme with geolocation database over TV white space," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3998–4007, May 2018.

[46] H. Sun, A. Nallanathan, C. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Communications Magazine*, vol. 20, no. 2, pp. 74–81, 2013.

[47] Y. Pei, Y. Liang, K. C. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, Sep. 2011.

[48] Z. Sun and J. N. Laneman, "Performance metrics, sampling schemes, and detection algorithms for wideband spectrum sensing," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5107–5118, 2014.

[49] Y. Pei, Y. Liang, K. C. Teh, and K. H. Li, "Energy-efficient design of sequential channel sensing in cognitive radio networks: Optimal sensing strategy, power allocation, and sensing order," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1648–1659, 2011.

[50] "RFeye 20-6 intelligent wideband receiver," Cambridge RF Systems, Tech. Rep., 2017. [Online]. Available: https://pages.crfs.com/hubfs/datasheets/node-20-6-datasheet.pdf [Accessed: 2020-05-31]

[51] "RFeye 100-18 intelligent wideband receiver," Cambridge RF Systems, Tech. Rep., 2017. [Online]. Available: https://pages.crfs.com/hubfs/datasheets/node-100-18-datasheet.pdf [Accessed: 2020-05-31]

[52] E. Hossain, D. Niyato, and Z. Han, *Introduction to cognitive radio.* Cambridge University Press, 2009, p. 39–74.

[53] X. Zhang, Y. Ma, Y. Gao, and S. Cui, "Real-time adaptively regularized compressive sensing in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1146–1157, Feb. 2018.

[54] H. Qi, X. Zhang, and Y. Gao, "Subspace-Aided Low-Complexity blind compressive spectrum sensing over TV whitespace," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018.

[55] Z. Han, H. Li, and W. Yin, "Sparse optimization algorithms," in *Compressive Sensing for Wireless Networks.* Cambridge: Cambridge University Press, 2013, ch. 4, pp. 69–117.

[56] E. J. 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, Dec. 2006.

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

[58] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2116–2125, Jun. 2012.

[59] X. Zhang, Z. Qin, and Y. Gao, "Dynamic adjustment of sparsity upper bound in wideband compressive spectrum sensing," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2014, pp. 1214–1218.

[60] X. Zhang, Y. Ma, Y. Gao, and W. Zhang, "Autonomous compressive-sensing-augmented spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 6970–6980, 2018.

[61] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

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

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

[64] Y. Wang, Z. Tian, and C. Feng, "A two-step compressed spectrum sensing scheme for wideband cognitive radios," in *IEEE Global Communications Conference (GLOBE-*

COM), 2010, pp. 1–5.

[65] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal on Selected Areas of Signal Processing*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[66] D. D. Arianandа, G. Leus, and Z. Tian, "Multi-coset sampling for power spectrum blind sensing," in *International Conference on Digital Signal Processing*, 2011, pp. 1–8.

[67] D. Cohen and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: A unified approach," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3897–3910, 2014.

[68] D. Cohen and Y. C. Eldar, "Sub-Nyquist cyclostationary detection for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 3004–3019, 2017.

[69] Z. Tian, Y. Tafesse, and B. M. Sadler, "Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing," *IEEE Journal on Selected Areas of Signal Processing*, vol. 6, no. 1, pp. 58–69, 2012.

[70] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1847–1862, 2010.

[71] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[72] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, "Analog-to-information conversion via random demodulation," in *IEEE Workshop on Design, Applications, Integration and Software*, 2006, pp. 71–74.

[73] M. A. Lexa, M. E. Davies, and J. S. Thompson, "Reconciling compressive sampling systems for spectrally sparse continuous-time signals," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 155–171, 2012.

[74] J. Yoo, S. Becker, M. Loh, M. Monge, E. Candès, and A. Emami-Neyestanak, "A

100MHz–2GHz 12.5x sub-Nyquist rate receiver in 90nm cmos," in *IEEE Radio Frequency Integrated Circuits Symposium*, 2012, pp. 31–34.

[75] T. Murray, P. Pouliquen, A. G. Andreou, and K. Lauritzen, "Design of a cmos a2i data converter: Theory, architecture and implementation," in *Annual Conference on Information Sciences and Systems (CISS)*, 2011, pp. 1–6.

[76] J. Yoo, S. Becker, M. Monge, M. Loh, E. Candès, and A. Emami-Neyestanak, "Design and implementation of a fully integrated compressed-sensing signal acquisition system," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 5325–5328.

[77] R. G. Baraniuk, "andom acquisition sensors and receivers: Design, architectures, and applications," Ph.D. dissertation, Rice University, Aug. 2017.

[78] S. R. Becker, "Practical compressed sensing: Modern data acquisition and signal processing," Ph.D. dissertation, California Institute of Technology, Jun. 2011.

[79] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 1996, pp. 1688–1691 vol. 3.

[80] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4692–4702, 2008.

[81] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits, Devices and Systems*, vol. 5, pp. 8–20, 2011.

[82] Y. Eldar, "Software and Hardware - Hardware - Xampling: MWC prototype." [Online]. Available: http://www.wisdom.weizmann.ac.il/~yonina/YoninaEldar/hardware_mwc.html [Accessed: 2020-05-31]

[83] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4628–4642, 2012.

[84] J. Treichler, M. Davenport, and R. Baraniuk, "Application of compressive sensing

to the design of wideband signal acquisition receivers," in *The 6th US and Australia Joint Workshop on Defense Applications of Signal Processing (DASP)*, Jan. 2009.

[85] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," Dec. 2019. [Online]. Available: https://friedlander.io/spgl1 [Accessed: 2020-05-31]

[86] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation (fpc) - an algorithm for large-scale image and data processing applications of l_1-minimization." [Online]. Available: https://www.caam.rice.edu/~optimization/L1/fpc [Accessed: 2020-05-31]

[87] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[88] X. Zhang, Y. Ma, H. Qi, Y. Gao, Z. Xie, Z. Xie, M. Zhang, X. Wang, G. Wei, and Z. Li, "Distributed compressive sensing augmented wideband spectrum sharing for cognitive iot," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3234–3245, 2018.

[89] X. Zhang, Y. Ma, H. Qi, and Y. Gao, "Low-complexity compressive spectrum sensing for large-scale real-time processing," *IEEE Wireless Communications Letters*, vol. 7, no. 4, pp. 674–677, 2018.

[90] M. E. Lopes, "Unknown sparsity in compressed sensing: Denoising and inference," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5145–5166, Sept. 2016.

[91] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Sept. 2001.

[92] Y. Gao, Y. Chen, and Y. Ma, "Sparse-Bayesian-learning-based wideband spectrum sensing with simplified modulated wideband converter," *IEEE Access*, vol. 6, pp. 6058–6070, 2018.

[93] D. Huang, S. Wu, W. Wu, and P. Wang, "Cooperative radio source positioning and power map reconstruction: A sparse Bayesian learning approach," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2318–2332, 2015.

[94] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE*

Transactions on Signal Processing, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[95] Y. Ma, X. Zhang, and Y. Gao, "Joint sub-nyquist spectrum sensing scheme with geolocation database over tv white space," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3998–4007, 2018.

[96] S. M. Kay, "Statistical detection theory I," in *Fundamentals of Statistical Signal Processing: Detection Theory.* Upper Saddle River: Prentice-Hall, 1993, ch. 3.

[97] W. Zhang, R. K. Mallik, and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.

[98] A. Ebrahimzadeh, M. Najimi, S. M. H. Andargoli, and A. Fallahi, "Sensor selection and optimal energy detection threshold for efficient cooperative spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1565–1577, 2015.

[99] R. Tandra and A. Sahai, "Snr walls for signal detection," *IEEE Journal on Selected Areas of Signal Processing*, vol. 2, no. 1, pp. 4–17, 2008.

[100] F. F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[101] W. A. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Transactions on Communications*, vol. 36, no. 8, pp. 897–906, 1988.

[102] A. V. Dandawate and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2355–2369, 1994.

[103] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2596–2605, Jul. 2012.

[104] M. Yang, Y. Li, X. Liu, and W. Tang, "Cyclostationary feature detection based spectrum sensing algorithm under complicated electromagnetic environment in cognitive radio networks," *China Communications*, vol. 12, no. 9, pp. 35–44, Sept.

2015.

[105] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[106] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, pp. 37–52, 1967.

[107] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer, 2010.

[108] P. Feng, "Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1998.

[109] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2173–2183, Sep 2000.

[110] R. G. Vaughan, N. L. Scott, and D. R. White, "The theory of bandpass sampling," *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 1973–1984, 1991.

[111] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429 – 443, 2011.

[112] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[113] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589 – 592, 2008.

[114] S. Foucart, "A note on guaranteed sparse recovery via l_1-minimization," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 97 – 103, 2010.

[115] Q. Mo and S. Li, "New bounds on the restricted isometry constant δ_{2k}," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 460 – 468, 2011.

[116] E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," vol. abs/1011.3854, 2010. [Online]. Available: http://arxiv.org/abs/1011.3854 [Accessed: 2020-05-31]

[117] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling,"

Inverse Problems, vol. 23, pp. 969–985, Jun. 2007.

[118] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 397–399, 1974.

[119] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, pp. 1030–1051, 2006.

[120] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Dec 2008.

[121] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[122] J. Haupt, L. Applebaum, and R. Nowak, "On the restricted isometry of deterministically subsampled fourier matrices," in *Annual Conference on Information Sciences and Systems (CISS)*, March 2010, pp. 1–6.

[123] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," in *Proceedings of the National Academy of Sciences*, vol. 100, Mar. 2003, p. 2197–2202.

[124] D. Cohen and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: A unified approach," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3897–3910, 2014.

[125] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Mathematical Research Letters*, vol. 12, no. 1, p. 121–127, 2005.

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

[127] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[128] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[129] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[130] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for l_1-minimization: Methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.

[131] D. Drusvyatskiy, "The proximal point method revisited," 2017. [Online]. Available: https://arxiv.org/pdf/1712.06038.pdf [Accessed: 2020-05-31]

[132] J. Nocedal, *Penalty and Augmented Lagrangian Methods.* New York, NY: Springer New York, 2006, pp. 497–528.

[133] M. S. Gockenbach, "The quadratic penalty method for equality-constrained optimization," Michigan Technical University, Tech. Rep., Mar. 2003. [Online]. Available: https://pages.mtu.edu/~msgocken/ma5630spring2003/lectures/pen/pen/pen.html [Accessed: 2020-05-31]

[134] D. P. Bertsekas, "Chapter 4 - exact penalty methods and Lagrangian methods," in *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982, pp. 179 – 301.

[135] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for l_1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.

[136] R. Tibshirani, I. Johnstone, T. Hastie, and B. Efron, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, p. 407–499, Apr 2004.

[137] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, 2005, pp. 733–736.

[138] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[139] J. Yang and Y. Zhang, "Alternating direction algorithms for l_1-problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.

[140] J. Eckstein, "Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results," *Rutcor Research Report (RRR) 32-2012*, Dec. 2012.

[141] M. Wang, W. Xu, and A. Tang, "On the performance of sparse recovery via ℓ_p-minimization $(0 \leq p \leq 1)$," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7255–7278, 2011.

[142] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[143] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.

[144] F. Cao, M. Cai, Y. Tan, and J. Zhao, "Image super-resolution via adaptive $\ell_p(0 < p < 1)$ regularization and sparse representation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, pp. 1550–1561, July 2016.

[145] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[146] T. Wimalajeewa and P. K. Varshney, "Omp based joint sparsity pattern recovery under communication constraints," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5059–5072, Oct. 2014.

[147] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, May. 2009.

[148] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[149] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, Dec. 2011.

[150] D. L. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[151] A. Roche, "EM algorithm and variants: an informal tutorial," 2011.

[152] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572 – 588, 2006, sparse Approximations in Signal and Image Processing.

[153] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589 – 602, 2006, sparse Approximations in Signal and Image Processing.

[154] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[155] H. Qi and Y. Gao, "On timing skews of multicoset samplers in compressive spectrum sensing for millimeter-wave," in *IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[156] M. Mishali and Y. C. Eldar, "Expected RIP: Conditioning of the modulated wideband converter," in *IEEE Information Theory Workshop*, 2009, pp. 343–347.

[157] J. C. Candy and G. Temes, *Oversampling Methods for A/D and D/A Conversion.* IEEE Press, 1992, ch. 1.

[158] Y. Arjoune and N. Kaabouch, "Wideband spectrum sensing: A bayesian compressive sensing approach," *Sensors*, vol. 18, p. 1839, Jun. 2018.

[159] A. Toma, T. Nawaz, Y. Gao, L. Marcenaro, and C. Regazzoni, "Interference mitigation in wideband radios using spectrum correlation and neural network," *IET Communications*, Feb. 2019.

[160] Z. Ye, G. Memik, and J. Grosspietsch, "Energy detection using estimated noise variance for spectrum sensing in cognitive radio networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2008, pp. 711–716.

[161] H. Qi, X. Zhang, and Y. Gao, "Channel energy statistics learning in compressive spectrum sensing," *IEEE Transactions on Wireless Communications*, vol. 17,

no. 12, pp. 7910–7921, Dec. 2018.

[162] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190–1196, Aug. 1989.

[163] Z. Lu and A. M. Zoubir, "Generalized bayesian information criterion for source enumeration in array processing," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1470–1480, Mar. 2013.

[164] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[165] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465 – 471, 1978.

[166] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[167] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.

[168] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, Jun 2007.

[169] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[170] Z. Lu and A. M. Zoubir, "Source enumeration using the pdf of sample eigenvalues via information theoretic criteria," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 3361–3364.

[171] T. W. Anderson, "Asymptotic theory for principal component analysis," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, 1963.

[172] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.

[173] A. Rakotomamonjy, "Algorithms for Multiple Basis Pursuit Denoising," in *Signal Processing with Adaptive Sparse Structured Representations*. Inria Rennes - Bretagne Atlantique, Apr. 2009. [Online]. Available: https://hal.inria.fr/inria-00369535 [Accessed: 2020-05-31]

[174] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *IEEE International Conference on Acoustics, Speech, and Signal Process. (ICASSP)*, vol. 5, Mar. 2005, pp. 721–724.

[175] M. O. Mughal, T. Nawaz, L. Marcenaro, and C. S. Regazzoni, "Cyclostationary-based jammer detection algorithm for wide-band radios using compressed sensing," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2015, pp. 280–284.

[176] D. A. Pados, P. Papantoni-Kazakos, D. Kazakos, and A. G. Koyiantis, "On-line threshold learning for neyman-pearson distributed detection," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 10, pp. 1519–1531, Oct. 1994.

[177] D. R. Joshi, D. C. Popescu, and O. A. Dobre, "Adaptive spectrum sensing with noise variance estimation for dynamic cognitive radio systems," in *Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2010, pp. 1–5.

[178] S. Gong, W. Liu, W. Yuan, W. Cheng, and S. Wang, "Threshold-learning in local spectrum sensing of cognitive radio," in *IEEE Vehicular Technology Conference (VTC) Spring*, Apr. 2009, pp. 1–6.

[179] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4388–4394, Sept. 2010.

[180] H. Qi, X. Zhang, and Y. Gao, "Mixture model-based channel energy statistics modeling in compressive spectrum sensing," in *IEEE International Conference on Communincations in China (ICCC)*, Aug. 2018.

[181] H. Qi and Y. Gao, "Two-dimensional compressive spectrum sensing in collaborative cognitive radio networks," in *IEEE Globe Conference on Communications (GLOBECOM)*, Dec. 2017.

[182] G. J. McLachlan and K. E. Basford, "Mixture models. Inference and applications

to clustering," in *Statistics: Textbooks and Monographs*. New York: Dekker, 1988.

[183] M. Z. Zheleva, R. Chandra, A. Chowdhery, P. Garnett, A. Gupta, A. Kapoor, and M. Valerio, "Enabling a nationwide radio frequency inventory using the spectrum observatory," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 362–375, Feb. 2018.

[184] M. Zheleva, R. Chandra, A. Chowdhery, A. Kapoor, and P. Garnett, "TxMiner: Identifying transmitters in real-world spectrum measurements," in *IEEE International Symposium on Dynamic Spectrum Access Network (DySPAN)*, Sept. 2015, pp. 94–105.

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

[186] T. Hastie, R. Tibshirani, and J. Friedman, "Model inference and averaging," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2013, ch. 8, pp. 236–242.

[187] N. Johnson, S. Kotz, and N. Balakrishnan, "Noncentral chi-square distributions," in *Continuous Univariate Distributions*. New York: Wiley, 1994, vol. 2.

[188] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, Mar. 1983.

[189] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[190] D. MacKay, "An example inference task: Clustering," in *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press, 2003, ch. 20.

[191] S. M. Kay, "Maximum likelihood estimation," in *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River: Prentice-Hall, 1993, ch. 7.

[192] J. M. Bernardo, "Algorithm as 103: Psi (digamma) function," *Applied Statistics*, vol. 25, no. 3, pp. 315–317, 1976.

[193] K. Burnham and D. Anderson, "Information and likelihood theory: A basis for model selection and inference," in *Model Selection and Multimodel Inference: A*

Practical Information-Theoretic Approach. New York: Springer, 2007, ch. 2.

[194] H. Johansson and P. Lowenborg, "Reconstruction of nonuniformly sampled bandlimited signals by means of digital fractional delay filters," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2757–2767, Nov. 2002.

[195] M. Hayes, J. Lim, and A. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 672–680, Dec. 1980.

[196] "Getting Started Guide - 2 GHz Bandwidth mmWave Transceiver System," National Instruments, Tech. Rep. [Online]. Available: http://www.ni.com/pdf/manuals/376531c.pdf [Accessed: 2020-05-31]

[197] "Specifications - PXIe-7902 high-speed serial FPGA module," National Instruments, Tech. Rep. [Online]. Available: http://www.ni.com/pdf/manuals/375758a.pdf [Accessed: 2020-05-31]

[198] M. Mishali and E. Yonina, "Spectrum-blind sampling and reconstruction of multiband signals," European Patent EP2188780B1, Jan. 2011.

[199] F. Marvasti and M. B. Mashadi, "Wideband analog to digital conversion by random or level crossing sampling," U.S. Patent US9729160B1, Jan. 2017.

[200] F. Lin and H. Deng, "Continuously tunable true-time-delay phase shifter based on transmission lines with simultaneously reconfigurable impedance and phase constant," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 12, pp. 4714–4723, 2019.

[201] C. Shafai, S. K. Sharma, L. Shafai, and D. D. Chrusch, "Microstrip phase shifter using ground-plane reconfiguration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 1, pp. 144–153, 2004.

[202] L. Martínez Cano, A. L. Borja, V. E. Boria, and A. Belenguer, "Highly versatile coplanar waveguide line with electronically reconfigurable bandwidth and propagation characteristics," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 1, pp. 128–135, 2017.

[203] L. Rabieirad and S. Mohammadi, "Reconfigurable cmos tuners for software-defined radio," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 11,

pp. 2768–2774, 2009.

[204] N. Kurosawa, H. Kobayashi, K. Maruyama, H. Sugawara, and K. Kobayashi, "Explicit analysis of channel mismatch effects in time-interleaved adc systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 3, pp. 261–271, 2001.

[205] D. M. Hummels, J. J. Mcdonald, and F. H. Irons, "Distortion compensation for time-interleaved analog to digital converters," in *IEEE Instrumentation and Measurement Technology Conference*, vol. 1, 1996, pp. 728–731 vol.1.

[206] Daihong Fu, K. C. Dyer, S. H. Lewis, and P. J. Hurst, "A digital background calibration technique for time-interleaved analog-to-digital converters," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, pp. 1904–1911, 1998.

[207] E. Iroaga, B. Murmann, and L. Nathawad, "A background correction technique for timing errors in time-interleaved analog-to-digital converters," in *IEEE International Symposium on Circuits and Systems*, 2005, pp. 5557–5560 Vol. 6.

[208] Z. Han, H. Li, and W. Yin, "Compressive sensing framework," in *Compressive Sensing for Wireless Networks*. Cambridge: Cambridge University Press, 2013, ch. 3, pp. 69–117.

Appendix A

Proof of Lemma 1 in Chapter 3

Proof. Denote the subspace dimension $\dim\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)}\} = v$, and naturally one has $v \leq s < r$. Consider the case that $\boldsymbol{\alpha}^{(s+1)}$ falls in the subspace $\text{span}\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)}\}$, where the probability equals to the integral of the probability density function that is absolutely continuous in space $\mathbb{C}^r$ on a v-dimension hyperplane, which has the measure of 0. Hence, with the probability of 1, $\boldsymbol{\alpha}^{(s+1)}$ is linear-independent with $\{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(s)}\}$, which completes the proof. □

Appendix B

Proof of Proposition 1 in Chapter 3

Proof. Given arbitrary $\boldsymbol{s}_f^{(1)}$, from Lemma 1, it holds that $\dim\{\boldsymbol{s}_f^{(1)}\ \boldsymbol{s}_f^{(2)}\} = 2$ with probability of 1. For $2 \leq s \leq \delta(\boldsymbol{s}_f) - 1$, assuming $\dim\{\boldsymbol{s}_f^{(1)}\ \cdots \boldsymbol{s}_f^{(s)}\} = s$, again from Lemma 1, it holds that $\dim\{\boldsymbol{s}_f^{(1)} \cdots \boldsymbol{s}_f^{(s+1)}\} = s+1$. Consequently, one can have the first $\delta(\boldsymbol{s}_f)$ samples satisfying $\dim\left\{\boldsymbol{s}_f^{(1)}\ \boldsymbol{s}_f^{(2)}\ \cdots, \boldsymbol{s}_f^{(\delta(\boldsymbol{s}_f))}\right\} = \text{rank}\left[\boldsymbol{s}_f^{(1)}\ \boldsymbol{s}_f^{(2)}\ \cdots, \boldsymbol{s}_f^{(\delta(\boldsymbol{s}_f))}\right] = \delta(\boldsymbol{s}_f)$. Then, with $P - \delta(\boldsymbol{s}_f)$ more samples as columns, one naturally has $\text{rank}\,(\boldsymbol{S}_f) \geq \delta(\boldsymbol{s}_f)$. On the other hand, consider the sparsity of $\boldsymbol{s}_f^{(p)}$'s, it also holds $\text{rank}\,(\boldsymbol{S}_f) \leq \delta(\boldsymbol{s}_f)$. From both inequalities, one has $\text{rank}\,(\boldsymbol{S}_f) = \delta(\boldsymbol{s}_f)$. From $\hat{\boldsymbol{R}}_{s_f} = \frac{1}{P}\boldsymbol{S_f}\boldsymbol{S_f}^H$, one can write $\text{rank}(\hat{\boldsymbol{R}}_{s_f}) = \text{rank}(\boldsymbol{S}_f) = \delta(\boldsymbol{s}_f)$.

Furthermore, consider $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}\boldsymbol{S}_f$, one can obtain $\text{rank}(\boldsymbol{Y}) = \min\left\{\text{rank}(\boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}), \delta(\boldsymbol{s}_f)\right\} = \delta(\boldsymbol{s}_f)$. Finally, it is reached that $\text{rank}(\hat{\boldsymbol{R}}_y) = \text{rank}(\frac{1}{P}\boldsymbol{Y}\boldsymbol{Y}^H) = \text{rank}(\boldsymbol{Y}) = \delta(\boldsymbol{s}_f)$. □

Appendix C

Proof of Proposition 2 in Chapter 3

Proof. $\forall \boldsymbol{x} \in \mathbb{C}^b$s.t.$\boldsymbol{M}^H\boldsymbol{x} = \boldsymbol{0}$, one has $\boldsymbol{M}\boldsymbol{M}^H\boldsymbol{x} = \boldsymbol{0}$, which means null$(\boldsymbol{M}^H) \subseteq$ null$(\boldsymbol{M}\boldsymbol{M}^H)$. Similarly, $\forall \boldsymbol{x} \in \mathbb{C}^b$ s.t. $\boldsymbol{M}\boldsymbol{M}^H\boldsymbol{x} = \boldsymbol{0}$, the quadratic form satisfies $\boldsymbol{x}^H\boldsymbol{M}\boldsymbol{M}^H\boldsymbol{x} = \boldsymbol{0}$, then $\boldsymbol{M}^H\boldsymbol{x} = \boldsymbol{0}$, which is equivalent to null$(\boldsymbol{M}\boldsymbol{M}^H) \subseteq$ null$(\boldsymbol{M}^H)$. Thus, it holds that null$(\boldsymbol{M}\boldsymbol{M}^H) =$ null$(\boldsymbol{M}^H)$. For matrix $\boldsymbol{N}$, again, one has null$(\boldsymbol{N}\boldsymbol{N}^H) =$ null$(\boldsymbol{N}^H)$ and it follows that null$(\boldsymbol{M}^H) =$ null$(\boldsymbol{N}^H) =$ null$(\boldsymbol{N}\boldsymbol{N}^H) =$ null$(\boldsymbol{M}\boldsymbol{M}^H)$. Finally, from rank-nullity theorem, it holds that span$(\boldsymbol{M}) =$ null$(\boldsymbol{M}^H)^\perp =$ span$(\boldsymbol{N}) =$ null$(\boldsymbol{N}^H)^\perp$, which completes the proof. □

Appendix D

Proof of Corollary 1 in Chapter 3

Proof. Consider the case $\boldsymbol{B}_v = \mathbf{0}$, and it is noted that $\boldsymbol{V}_s \boldsymbol{\lambda}_s \boldsymbol{V}_s^H = \boldsymbol{Y}'_s \boldsymbol{Y}'^H_s$ when $k^c = \delta(\boldsymbol{s}_f)$. From Proposition 2 one has span $\left(\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}\right) = \text{span}(\boldsymbol{Y}_s)$. Since columns of the two matrices span the same k^c-dimension subspace in $\mathbb{C}^M$, they can be linked via a unique linear transform characterized by full-(column-)rank matrix $\boldsymbol{L}_{P\times k^c}$, that is $\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}} = \boldsymbol{Y}_s\boldsymbol{L}$. There exists a matrix $\boldsymbol{S}_v = \boldsymbol{S_f L}$ that satisfies $\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}} = \boldsymbol{\Phi S}_v = \boldsymbol{Y}_s\boldsymbol{L} = \boldsymbol{\Phi S}_f\boldsymbol{L}$. And such $\boldsymbol{S}_v$ apparently has the same support as $\boldsymbol{S}_f$ (i.e. equality in (3.24)) with row sparsity of $\delta(\boldsymbol{s}_f)$.

For underestimated dimensionality of signal subspace $k^c < \delta(\boldsymbol{s}_f)$, the detected signal subspace has reduced dimensionality compared to that of $\boldsymbol{Y}_s$, that is span $\left(\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}\right) \subset$ span$(\boldsymbol{Y}_s)$. Then, the relationships $\boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}} = \boldsymbol{Y}_s\boldsymbol{L}$ and $\boldsymbol{S}_v = \boldsymbol{S}_f\boldsymbol{L}$ hold where there exists a matrix $\boldsymbol{L}_{P\times\delta(\boldsymbol{s}_f)}$ with the rank of k^c. Similarly, this leads to supp$(\boldsymbol{S}_v) \subseteq$ supp$(\boldsymbol{S}_f)$. □

Appendix E

Proof for Proposition 3 in Chapter 3

Proof. Before formally starting the proof, the following Lemma 2 is introduced directly from the definition of restricted isotropic property (RIP).

Lemma 2 ([149])**.** *Define the RIP constant of matrix* $\boldsymbol{A}$ *with sparsity* s *as*

$$\sigma_{\boldsymbol{A},s} := \max_{\{S|\,\mathrm{card}(S)\leq s\},\,\forall \boldsymbol{x}\in\mathbb{C}^S} \frac{\left\langle \left((\boldsymbol{A})_{:,S}^H (\boldsymbol{A})_{:,S} - \boldsymbol{I} \right) \boldsymbol{x}, \boldsymbol{x} \right\rangle}{||\boldsymbol{x}||_2^2}, \tag{E.1}$$

and the following relations hold

$$\begin{aligned} \left|\left| \left\langle (\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A})\boldsymbol{u}, \boldsymbol{v} \right\rangle \right|\right|_2 &\leq \sigma_t ||\boldsymbol{u}||_2 ||\boldsymbol{v}||_2, \\ &\forall \boldsymbol{u}, \boldsymbol{v} \text{ s. t. } \mathrm{card}\left(\mathrm{supp}(\boldsymbol{u}) \cup \mathrm{supp}(\boldsymbol{v})\right) \leq t, \end{aligned} \tag{E.2}$$

and

$$\begin{aligned} \left|\left| \left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A} \right) \boldsymbol{v} \right)_U \right|\right|_2 &\leq \sigma_t ||\boldsymbol{v}||_2, \\ &\forall U, \boldsymbol{v} \text{ s. t. } \mathrm{card}\left(U \cup \mathrm{supp}(\boldsymbol{v})\right) \leq t. \end{aligned} \tag{E.3}$$

For conciseness, denote $\boldsymbol{A} := \boldsymbol{\Phi}\boldsymbol{\mathcal{F}}^{-1}$ and $\boldsymbol{M}_s := \boldsymbol{V}_s(\boldsymbol{\Lambda}_s)^{\frac{1}{2}}$. From the least-square step (line 4) in Algorithm 1, the residue produced in the $(i+1)$th iteration $\boldsymbol{A}\boldsymbol{S}_v^{(i+1)} - \boldsymbol{M}_s$ is orthogonal to the measurement space determined by the selected block support, span $\left((\boldsymbol{A})_{\mathrm{BLK},\Omega^{(i+1)}}\right)$. This leads to

$$(\boldsymbol{A}\boldsymbol{S}_{\boldsymbol{v}})^H \left(\boldsymbol{A}\boldsymbol{S}_v^{(i+1)} - \boldsymbol{M}_s\right) = \boldsymbol{0}, \forall \boldsymbol{S}_v \text{ s. t. } (\boldsymbol{S}_v)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}} = \boldsymbol{0}, \tag{E.4}$$

which can be rewritten as

$$\left(\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{S}_{\boldsymbol{v}}\right)^H \left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right) = \boldsymbol{0}, \forall \boldsymbol{S}_v \text{ s. t. } (\boldsymbol{S}_v)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}} = \boldsymbol{0}. \tag{E.5}$$

Then, examine the error from the detected support

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right|\right|_F^2 \\
=& \operatorname{tr}\left(\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)^H \left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right) \\
\underset{(E.5)}{=}& \operatorname{tr}\Bigg(\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)^H \\
&\qquad \cdot \left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\Bigg) \\
\underset{(E.2)}{\leq}& \sigma_{\boldsymbol{A},2k^c}\left|\left|\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right|\right|_F \left|\left|\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right|\right|_F,
\end{aligned} \tag{E.6}$$

from which one can arrive at

$$\left|\left|\left(\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right|\right|_F \leq \sigma_{\boldsymbol{A},2k^c}\left|\left|\boldsymbol{S}_v^{(i+1)} - \boldsymbol{S}_v\right|\right|_F. \tag{E.7}$$

Next, the total error at the $(i+1)$th iteration is accounted,

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)\right|\right|_F^2\\
=&\left|\left|\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right|\right|_F^2\\
&+\left|\left|\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F^2\\
\leq&\sigma^2_{\boldsymbol{A},2k^c}\left|\left|\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right|\right|_F^2+\left|\left|\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F^2,
\end{aligned}\tag{E.8}$$

from which one can obtain

$$\begin{aligned}
&\left|\left|\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right|\right|_F^2\\
\leq&\frac{1}{1-\sigma^2_{\boldsymbol{A},2k^c}}\left|\left|\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F^2.
\end{aligned}\tag{E.9}$$

On the other hand, examine the correlation step (line 3) in Algorithm 1 and one instantly has the following

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v^{(i)}+\boldsymbol{A}^H\left(\boldsymbol{M}_s-\boldsymbol{A}\boldsymbol{S}_v^{(i)}\right)\right)_{\mathrm{BLK},\Omega^{(i+1)}}\right|\right|_F^2\\
\geq&\left|\left|\left(\boldsymbol{S}_v^{(i)}+\boldsymbol{A}^H\left(\boldsymbol{M}_s-\boldsymbol{A}\boldsymbol{S}_v^{(i)}\right)\right)_{\mathrm{BLK},\Omega_s}\right|\right|_F^2,
\end{aligned}\tag{E.10}$$

where Ω_s is the block support of $\boldsymbol{S}_v$. The above may be rewritten as

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v^{(i)}+\boldsymbol{A}^H\left(\boldsymbol{M}_s-\boldsymbol{A}\boldsymbol{S}_v^{(i)}\right)\right)_{\mathrm{BLK},\Omega^{(i+1)}\setminus\Omega_s}\right|\right|_F^2\\
=&\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\Omega^{(i+1)}\setminus\Omega_s}\right|\right|_F^2\\
\geq&\left|\left|\left(\boldsymbol{S}_v^{(i)}+\boldsymbol{A}^H\left(\boldsymbol{M}_s-\boldsymbol{A}\boldsymbol{S}_v^{(i)}\right)\right)_{\mathrm{BLK},\Omega_s\setminus\Omega^{(i)}}\right|\right|_F^2.
\end{aligned}\tag{E.11}$$

The right-hand side of the inequality (E.11) is reorganized as the following given the fact

$$\left(\boldsymbol{S}_v^{(i+1)}\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}=\boldsymbol{0},$$

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v^{(i)}+\boldsymbol{A}^H\left(\boldsymbol{M}_s-\boldsymbol{A}\boldsymbol{S}_v^{(i)}\right)\right)_{\mathrm{BLK},\Omega_s\setminus\Omega^{(i+1)}}\right|\right|_F\\
=&\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\Omega_s\setminus\Omega^{(i+1)}}\right.\right.\\
&\left.\left.+\left(\boldsymbol{S}_v-\boldsymbol{S}_v^{(i+1)}\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F\\
\geq&\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\Omega_s\setminus\Omega^{(i+1)}}\right|\right|_F\\
&+\left|\left|\left(\boldsymbol{S}_v-\boldsymbol{S}_v^{(i+1)}\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F.
\end{aligned}\tag{E.12}$$

Focus on the last term and one can shrink it from (E.12) and (E.11)

$$\begin{aligned}
&\left|\left|\left(\boldsymbol{S}_v-\boldsymbol{S}_v^{(i+1)}\right)_{\mathrm{BLK},\overline{\Omega^{(i+1)}}}\right|\right|_F\\
\leq&\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\Omega_s\setminus\Omega^{(i+1)}}\right|\right|_F\\
&+\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\Omega^{(i+1)}\setminus\Omega_s}\right|\right|_F\\
\leq&\sqrt{2}\left|\left|\left(\left(\boldsymbol{I}-\boldsymbol{A}^H\boldsymbol{A}\right)\right.\right.\right.\\
&\left.\left.\left.\cdot\left(\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right)\right)_{\mathrm{BLK},(\Omega_s\setminus\Omega^{(i+1)})\cup(\Omega^{(i+1)}\setminus\Omega_s)}\right|\right|_F\\
\underset{(E.3)}{\leq}&\sqrt{2}\sigma_{\boldsymbol{A},3k^c}\left|\left|\left(\boldsymbol{S}_v-\boldsymbol{S}_v^{(i)}\right)\right|\right|_F.
\end{aligned}\tag{E.13}$$

Congregate (E.9) and (E.13) and one can finally obtain the recurrence relation of error, writing

$$\left|\left|\boldsymbol{S}_v^{(i+1)}-\boldsymbol{S}_v\right|\right|_F\leq\sqrt{\frac{2\sigma_{\boldsymbol{A},3k^c}^2}{1-\sigma_{\boldsymbol{A},2k^c}^2}}\left|\left|\boldsymbol{S}_v^{(i)}-\boldsymbol{S}_v\right|\right|_F.\tag{E.14}$$

Convergence of the algorithm requires the constant in (E.14) less than one. By definition, $\sigma_{\boldsymbol{A},3k^c}>\sigma_{\boldsymbol{A},2k^c}$ naturally holds. This immediately leads to $\sigma_{\boldsymbol{A},3k^c}<1/\sqrt{3}$ as a guarantee of JB-HTP's convergence, which completes the proof. □

Appendix F

Proof of Proposition 4 in Chapter 3

Proof. The final iteration of JB-HTP is determined at the point where $\Omega^{(i-1)} = \Omega^{(i)} = \Omega = \Omega_s$. For starter, consider the selected block upon convergence $\forall p \in \Omega$ and others $\forall q \in \overline{\Omega}$, and the following can be obtained

$$\begin{aligned} &\left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\mathrm{BLK},\{p\}}\right|\right|_F \\ &> \left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\mathrm{BLK},\{q\}}\right|\right|_F \end{aligned} \tag{F.1}$$

On the left-hand side, it can be observed

$$\begin{aligned} &\left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\mathrm{BLK},\{p\}}\right|\right|_F \\ =&\left|\left|\left(\boldsymbol{S}_v\right)_{\mathrm{BLK},\{p\}} + \left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\{p\}}\right|\right|_F \\ \geq&\epsilon - \left|\left|\left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\mathrm{BLK},\{p\}}\right|\right|_F, \end{aligned} \tag{F.2}$$

where $\epsilon := \min_{p\in\Omega} \left\|(\boldsymbol{S}_v)_{\text{BLK},\{p\}}\right\|_F$. On the right-hand side, given the fact that $(\boldsymbol{S}_v)_{\text{BLK},\{q\}} = \boldsymbol{0}$, it holds that

$$\begin{aligned} &\left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\text{BLK},\{q\}}\right|\right|_F \\ =&\left|\left|\left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\text{BLK},\{q\}}\right|\right|_F. \end{aligned} \tag{F.3}$$

Subtract (F.3) from (F.2), and one may write and further re-organize as follows,

$$\begin{aligned} &\left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\text{BLK},\{p\}}\right|\right|_F \\ &- \left|\left|\left(\boldsymbol{S}_v^{(i-1)} + \boldsymbol{A}^H\left(\boldsymbol{M}_s - \boldsymbol{A}\boldsymbol{S}_v^{(i-1)}\right)\right)_{\text{BLK},\{q\}}\right|\right|_F \\ \geq&\epsilon - \left|\left|\left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\text{BLK},\{p\}}\right|\right|_F \\ &- \left|\left|\left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\text{BLK},\{q\}}\right|\right|_F \\ \underset{(E.2)}{\geq}&\epsilon - \sqrt{2}\left|\left|\left(\left(\boldsymbol{I} - \boldsymbol{A}^H\boldsymbol{A}\right)\left(\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right)\right)_{\text{BLK},\{p,q\}}\right|\right|_F \\ \geq&\epsilon - \sqrt{2}\sigma_{\boldsymbol{A},3k^c}\left\|\boldsymbol{S}_v^{(i-1)} - \boldsymbol{S}_v\right\|_F \\ \underset{(E.14)}{\geq}&\epsilon - \sqrt{2}\sigma_{\boldsymbol{A},3k^c}\left(\sqrt{\frac{2\sigma_{\boldsymbol{A},3k^c}^2}{1-\sigma_{\boldsymbol{A},2k^c}^2}}\right)^{i-1}\left\|\boldsymbol{S}_v^{(0)} - \boldsymbol{S}_v\right\|_F \\ =&\epsilon - \sqrt{1-\sigma_{\boldsymbol{A},2k^c}^2}\cdot\mu^i\left\|\boldsymbol{S}_v^{(0)} - \boldsymbol{S}_v\right\|_F \\ \underset{\text{Prop. 3}}{>}&\epsilon - \sqrt{\frac{2}{3}}\cdot\mu^i\left\|\boldsymbol{S}_v^{(0)} - \boldsymbol{S}_v\right\|_F, \end{aligned} \tag{F.4}$$

where $\mu := \sigma_{\boldsymbol{A},3k^c}\Big/\sqrt{1-\sigma_{\boldsymbol{A},2k^c}^2}$. To guarantee (F.2) to hold, inequality (F.4) may be bounded as following where i can be determined as a conservative upper bound,

$$\begin{aligned} 0 <&\epsilon - \sqrt{\frac{2}{3}}\cdot\mu^i\left\|\boldsymbol{S}_v^{(0)} - \boldsymbol{S}_v\right\|_F \\ i <&\left\lceil\frac{\log\left(\sqrt{\frac{2}{3}}\cdot\frac{\left\|\boldsymbol{S}_v^{(0)} - \boldsymbol{S}_v\right\|_F}{\epsilon}\right)}{\log\frac{1}{\mu}}\right\rceil. \end{aligned} \tag{F.5}$$

The specific upper bound of iteration number relies on the original sparse signal $\boldsymbol{S}_v$ which is a statistical signal, and the initial value $\boldsymbol{S}_v^{(0)}$. Here, consider a common case where $\boldsymbol{S}_v^{(0)} = \mathbf{0}$ and use a constant ρ to denote the expected logarithm ratio between the average received energy among all active channels and the minimum received channel energy,

$$\rho := \mathrm{E}\left[\log\left(\frac{\|\boldsymbol{S}_v\|_F}{l\epsilon}\right)\right]. \tag{F.6}$$

Get expectations on both sides (F.5) and obtain

$$\begin{aligned}
\mathrm{E}[i] &< \mathrm{E}\left[\left\lceil \frac{\log\left(\sqrt{\frac{2}{3}}\cdot\frac{\|\boldsymbol{S}_v\|_F}{\epsilon}\right)}{\log\frac{1}{\mu}} \right\rceil\right] \\
&\leq \mathrm{E}\left[\frac{\log\left(\sqrt{\frac{2}{3}}\cdot\frac{\|\boldsymbol{S}_v\|_F}{\epsilon}\right)}{\log\frac{1}{\mu}}\right] + 1 \\
&= \frac{\log\left(\sqrt{\frac{2}{3}}\right) + \log l + \mathrm{E}\left[\log\left(\frac{\|\boldsymbol{S}_v\|_F}{l\epsilon}\right)\right]}{\log\frac{1}{\mu}} + 1 \\
&= \frac{\log\left(\sqrt{\frac{2}{3}}\right) + \log l + \rho}{\log\frac{1}{\mu}} + 1 = \mathcal{O}(\log l) + \mathcal{O}(1),
\end{aligned} \tag{F.7}$$

which completes the proof. □