SPARSE SAMPLING AND ARRAY FOR WIRELESS COMMUNICATIONS

by

QIONG WU

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2015

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Qilian Liang, for all his guidance and encouragement. I have a wonderful experience in my years as a Ph.D. student largely because of him. He has made himself available over all these years whenever I want to have a discussion. I want to thank him for all the wise advices he has given to me on research, career and life. His enthusiasm and his taste in research problems have deeply influenced me and helped me to develop as a better researcher.

I also want to thank the faculty members in Electrical Engineering at UTA for the friendly, collaborative and stimulating atmosphere they have created for the department. In particular, I want to express my gratitude to Dr. Ioannis Schizas, Dr. Jean Gao, Dr. Jonathan Bredow, Dr. Saibun Tjuatja for serving on my comprehensive exam committee.

I feel in debt to all my colleagues and visiting scholars in the Wireless Communication Lab: Xin Wang, Junjie Chen, Zhuo Li, Ishrat Maherin, Shitong Yuan, Na Wu, Dr. Zongjie Cao, Dr. Feng Zhao, and Dr. Rui Min. Thanks for all the things they taught me! Special thanks to my friends and role models Yudong Chen at UC Berkeley and Yibo Jiao at Amazon for helping me out in my toughest time. I deeply enjoyed the time we spent discussing, arguing, and pair programming together!

I want to express my sincere thanks to Guozhu Wen and Jason Zhao at Douban, Cornelius Van Rensburg and Peter Wang at Huawei, and Rowan Chakoumakos, Amos Schallich, and Nick Weaver at EERO, for those enjoyable internships. These experiences have greatly broadened my research horizon.

Last but not least, I am immensely grateful to my parents. I want to thank them the most for always being supportive for all my choices, and offering wise advices and selfless helps through my toughest time. I have no way to thank them for their priceless love, care, inspiration, patience, and encouragement, but to dedicate this thesis to them.

April 7, 2015

ABSTRACT

SPARSE SAMPLING AND ARRAY FOR WIRELESS COMMUNICATIONS

Qiong Wu, Ph.D.

The University of Texas at Arlington, 2015

Supervising Professor: Qilian Liang

In recent years, sparse sampling and array, such as coprime sampling, nested sampling, nested array, attract attention for their potential to estimate autocorrelation coefficients with all lags, and to calculate the power spectrum density. But this theoretical merit is based on the premise that the input signals are wide-sense stationary. This dissertation first implements the coprime sampling for non-stationary signal, which still decreases the computational complexity but suffers from high variances due to lack of observations. Then, this algorithm is extended to calculate ambiguity function in the radar system, and several practical guidelines are concluded in order to conduct the sparse sensing while retain the detection quality.

The dissertation also proposes an algorithm deriving higher-order statistics (HOS) from the co-prime sampling and implements it in the order determination. Specifically, it is extended to pairwise co-prime sequences (PCS) for calculating HOS, and singular value decomposition (SVD) is used to the matrix formed by third-order cumulants to determine the order of the autoregressive model. In this scenario, PCS-based HOS algorithm enlarges the variance of third-order cumulants, which provides better indicator to determine the order of moving average (MA) model.

Furthermore, the dissertation extended the derivation of HOS based on nested sampling. It develops multilevel nested sampling (MNS) algorithm to obtain higher-order statistics (HOS), and analyzes the computational complexity of the MNS-HOS algorithm for both parametric and nonparametric methods. Compared to the existing HOS algorithms, the proposed algorithm vastly reduces the complexity by several orders in terms of the length of segmentation window. It also applies MNS-HOS algorithm to estimate the coefficients of a simplified LTE spatial channel model blindly without using any training sequences.

Moreover, the dissertation proposes a novel deployment for multi-cell cooperative cellular networks based on the two-dimensional (2D) nested co-array, and analyzes its sum-rate capacity and spectrum efficiency. It takes advantage of the invariance in the difference co-array so that the 2D nested array is able to calculate all elements in the covariance matrix of channel fading coefficients. Based on this premise, it demonstrates that the derivation procedure of average sum-rate capacity for the co-operative cellular networks is still valid for the nested distributed base stations (BSs) in the non-fading and Rayleigh fading channels. Given the same number of BSs, the proposed distribution significantly increases the sum-rate capacity of the system.

In addition, the dissertation also derives nested deployment for the hybrid wireless networks, and analyzes its ergodic throughput capacity in Rayleigh fading channels. As opposed to the existing hybrid wireless networks in which BS only serves individual cell area, the model of nested distributed BSs can be regarded as a multi-cell virtual multiple-input and multiple-output (MIMO) system. Meanwhile, the 2D nested deployment maximizes the degrees of freedom offered by the existing BSs from $O(N)$ to $O(N^2)$, in which the N stands for the number of BSs within the system. The ergodic capacity is investigated as an indicator of the system performances.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

Introduction

The degree of freedom (DoF) of sampling defined the minimum number of sample points, which could specify certain properties of the sequence as a whole [2]. Before the research of coprime samplers, the available sensors were considered as a signal array and increasing DoF could be achieved by performing an augmentation algorithm on the covariances obtained via minimum redundancy arrays (MRA) [3], which consisted uniformly linear arrays with maximum possible aperture. Bedrosian [4] extended the linear array to non-uniformly distribution such that their pairwise differences could generate full coverage for certain span, the paper also enumerated the array size M from 3 to 11 to achieve full coverage as much as $M(M-1)/2$. The algorithm proposed in [5] could find near-optimal integer sensor locations that maximized the number of distinct nonnegative integers, but it also restated the fact that location of elements in an MRA could only be approximated rather than specified in closed form. Besides, there were other ways to generate extra freedoms, including higher order statistics based methods, Khatri-Rao product based methods, and nested array [6]. Besides, the paper [8] developed the application of nested array beyond focusing on the DoF, finding nested array could improve the spectrum efficiency.

Coprime sampling first had been used for identifying sinusoids in noise [26] along with other methods proposed for synthetic aperture radar locating and imaging of moving targets [9]. Further research explored the properties and applications of coprime sampling and array in both time and frequency domains. The paper [27] used coprime samplers to increase the dimensions of DFT filter banks after sensor arrays

as well as to estimate the power spectrum density of received signal. In the paper [29], the multidimensional coprime sensing extended the previous implementations to acquire densely sampled domain. The paper [7] proposed spatial smoothing algorithm together with coprime sampling to estimate frequencies of sinusoids buried in noise and directions-of-arrival of impinging signals on a sensor array.

In communication, higher-order statistics (HOS) approaches are favorable to preserve both phase and amplitude information of the signal, and to deal with non-Gaussian sequences and nonlinear non-minimum phase system. Its estimates can be calculated from either conventional nonparametric methods of Fourier transform [23], or parametric methods based on moving average (MA), autoregressive (AR), or autoregressive moving average (ARMA) models [69]. The paper [23] provides further details about both kins of methods. It is worth noting that both of them have high variance and require a large number of records to obtain smooth estimates [24], but increasing the number of segments is demanding on computation, and may increase bias and introduce non-stationarity. Although there are preliminary researches [69,70] providing a sparse sampling scheme, pairwise coprime sequences (PCS), as a possible solution, how to accelerate speed of convergence and to leverage the trade-off between complexity and performance are still open questions.

Besides the implementation of signal processing, the nested distributed array is also introduced for signal acquisition via sensor array. It is firstly introduced to perform array processing with increased degrees of freedom using much fewer physical sensors [6]. The paper [73] generalizes this concept to the multiple dimensions, and provides the optimal structure to maximize the number of elements in the virtual co-array, as well as derives closed-form expressions for the sensor locations and the exact degrees of freedom obtainable from the proposed array as a function of the total number of the sensors.

Chapter 2 implements the coprimes sampling theory to process the non-stationary signal. Chapter 3 and 4 is to extend the use of co-prime sampling to higher-order statistics, which decreases the computational overhead while still retains all merits of cumulant-based order selection. In Chapter 5, the nested sampling scheme is further derived for HOS, and compared with the algorithm proposed in the previous chapter. In the Chapter 6 and 7, a novel nested-distributed network is introduced to advance the research on joint multi-cell processing, which is the first literature analyzing the communication systems with the nested-distributed BSs from the information-theoretic point of view.

CHAPTER 2

Coprime sampling for non-stationary signal processing

2.1 Introduction

Both of the designs of radar system and sensor network could be attribute to obtaining sufficient samples to generate the correlation function so that a good ambiguity scale or spectrum estimation could be obtained [1]. The design of radar system needs to take advantage of the ambiguity function (AF) between received signal and transmitted signal to determine the resolution of the radar, side lobe behavior, and ambiguities in both time and Doppler domains. AF is calculated via the convolution of transmitted signal with received signal, which contains the copy of transmitted signal, noise, and Doppler shift caused by the movement of the target. Furthermore, considering cost of deployment in broad range, many applications of sensor network require to distribute the sensor elements sparsely. The power spectral density (PSD) acquired by these sensors could describe the power incidents for the given direction and area, and PSD is the Fourier transform of autocorrelation function of received signal or correlation function among the signal received in different sensors in the array. Hence, both scenarios could benefit from sparse sensing a rapidly changing signal sequence with optimal performance in terms of retaining the resolution or detecting ability compared with dense sampling.

Note that the paper presenting coprime sampling [27] strictly confines discussion within the underlying assumption of wide-sense stationary signal so that the expectation of autocorrelation could approach the real value via multi-times averaging. This increased delay is used to compensate the variation introduced by sub-Nyquist

sampling. On the other hand, however, in the real-world application, just as the description in the first paragraph, the working scenarios of many applications involves non-stationary signal. The sampled points could not simply ascribe to independent and identical distribution either. Consequently, the autocorrelation coefficients might change dramatically during a short period. In Chapter **??**, we deal with this inconsistency and discuss the coprime sampling for non-stationary signal to obtain its second order statistic properties. In general, the classic point of view for processing non-stationary signal regards it as piece-wise stationary signal, but as these two theories combining together there are many research problems such as stability of estimation, coverage of second order derivatives, and so on. In the following content, we will discuss these problems and our tentative solutions in detail.

The rest of Chapter 2 is organized as follows, we first quickly revisit the basic concepts and properties about coprime sampling in Section 2.2. In Section 2.3, we propose and simulate the algorithm of two-steps coprime sampling especially used for the non-stationary signal. In Section 2.4, we extend the implementation scenario to radar signal processing and discuss several critical trade-offs in designing the radar signal processing system with coprime sampling. Finally, we conclude the research discussion in Section 2.5.

2.2 Theory and properties for coprime sampling

The algorithm of coprime sampling was introduced in [27]. The input signal is $S(T)$. Original sample rate is T_s, and the down sampling rate for two sample streams are M and N whose greatest common divisor is one. Then, except the beginning

point, the two generated sample streams do not have any overlap in origin signal sequence.

$$\begin{aligned} x_1[n_1] &= S(MT_s) \\ x_2[n_2] &= S(NT_s) \end{aligned} \tag{2.1}$$

Definition 1. *The difference co-array $x_k[n_1, n_2]$ is generated by two sample sequences $x_1[n_1]$ and $x_2[n_2]$ coprime sampled from input signal. Its index k satisfies*

$$k = Nn_1 - Mn_2 \quad -n_1 \in [0, \lfloor \frac{L}{M} \rfloor],\ n_2 \in [0, \lfloor \frac{L}{N} \rfloor] \tag{2.2}$$

The markers $\lfloor Z \rfloor$ stand for the largest integer less than certain values Z, and L stands for the total length of the signal segment. The coprimality of M and N can be used to show that the range of distinct value in $x_k[n_1, n_2]$ is the product of the coprime factors [27]. That is

$$-MN + 1 \leq Nn_1 - Mn_2 \leq MN - 1 \tag{2.3}$$

First of all, the physical meaning of this difference co-array is that via this difference co-array between the two coprime sampled steams the correlation of the original sequence could be calculated at all lags. Note that it does not confine the rate of down sampling, which might result the sample rate way below the Nyquist-sampling restriction. That is, the sampling might be arbitrarily sparse. On the other hand, however, there are two major drawbacks relevant with large values of coprime pairs: the latency in the time domain and the resolution range in the frequency domain. We will discuss them in detail in the following Section.

Besides, the minor differences in value ranges of coprime sampled signal streams generate different coverage of difference co-array and result in different coverage of autocorrelation coefficients.

Property 1. *With n_1 and n_2 restricted to the range $0 \leq n_1 \leq N - 1$ and $0 \leq n_2 \leq M - 1$, index of the resulting difference co-array $k = Mn_1 - Nn_2$ will have MN*

distinct values in the range $-(M-1)N \leq k \leq (N-1)M$*, which also indicates that there are absent values in the given range of* k*.*

Property 2. *If the ranges of* n_1 *and* n_2 *are* $0 \leq n_1 \leq N-1$ *and* $-M+1 \leq n_2 \leq M-1$*, the resulting index of difference co-array will achieve full coverage for* $0 \leq k \leq MN-1$*.*

The detailed demonstration of two properties above could be found in [27]. Furthermore, in this Chapter, we implement coprime sampling beyond the limit of $MN-1$, which leads to the following property.

Property 3. *Given sample points in the range* $(-L, L)$*, the largest coprime pair that it could have is* M *and* N *subject to* $MN < L$*, such that* n_1 *and* n_2 *restricted to the range* $0 \leq n_1 \leq \lfloor L/M \rfloor$ *and* $-\lfloor L/N \rfloor \leq n_2 \leq \lfloor L/N \rfloor$*, the resulting index of difference co-array* $k = Mn_1 - Nn_2$ *will achieve full coverage in the range* $0 \leq k \leq L-1$*.*

Proof: Following from the Euclid's Theorem [16], we could conclude that with any integer k in the range $[0, L-1]$, there are always integers $n_1{}'$ and $n_2{}'$ such that $k = Mn_1{}' - Nn_2{}'$.

Adding lMN to both terms in the right hand side of the formula with proper selection of variable l, we could let $n_1 = n_1{}' + lN$ such that $n_1 \in [0, \lfloor L/M \rfloor]$. Then we have

$$k = M(n_1{}' + lN) - N(n_2{}' + lM) \tag{2.4}$$

$$N(n_2{}' + lM) = M(n_1{}' + lN) - K \tag{2.5}$$

Since we have already known that $k \in [0, L-1]$ and $Mn_1 \in [0, L]$, the range of $N(n_2{}'+lM)$ should be $[-L, L]$. Let $n_2 = n_2{}'+lM$, we could have $n_2 \in [-\lfloor L/N \rfloor \lfloor L/n \rfloor]$ which concludes the proof.

Moreover, in the range $-MN + 1 \leq k \leq 0$, there are still absent values. But based on the symmetry property of autocorrelation, these results could be used for averaging the expectation of the symmetric positive counterpart.

2.3 STFT for coprime sampling non-stationary signal

2.3.1 Short time Fourier transform with coprime sampling

The presumption to generate autocorrelation from the coprime sampled sequence based on the previous chapter is that the second-order expectations of the sequence remain unchanged over time, which is essentially the wide-sense stationary (WSS) signal. In the application of radar signal processing, however, this criteria cannot hold anymore. In this Section, we will discuss how to combine coprime sampling with short time Fourier transform (STFT-CS) to process non-stationary signal, and demonstrate this algorithm is useful to preserve both the original quality of the signal and at the same time dramatic decrease the sample rate.

The choice of Short Time Fourier transform (STFT) is because this method is widely used in analyzing the time-frequency properties of non-stationary signals. In an STFT, the signal is segmented by a window function and performed Fourier transform within the window. The width of the window is a trade-off between temporal resolution and frequency resolution–better time resolution is achieved by narrow window while wider window could achieve better frequency resolution. In addition, in the scenario of coprime sampling, based on the *Property 3*, the window size also dictates the upper bound of the values of coprime pairs. Consequently, it determines the trade-off between stability of the estimation and the computational complexity of STFT-CS.

First of all, there is one definition to simplify the description of algorithm. Because the number of available autocorrelation estimation is changing along with the choice of coprime pairs, we define the procedure of finding the average as a single operator.

Definition 2. *$E(R_{xy}(k))$ stands for mathematical expectation of autocorrelation $R(k)$ for a given k using all available estimations. The value of k is determined by two independent index variables of the input sequence x and y.*

The algorithm involves several important independent variables listed in Table 1.

Based on STFT, within every slicing window we consider the sequence

$$x[n] = \begin{cases} s[n] & 0 \leq n \leq L-1, \\ 0 & otherwise, \end{cases} \tag{2.6}$$

The estimate of autocorrelation is

$$\hat{\varphi}_{xx}[m] = E(c_{xx}[m]) \tag{2.7}$$

where $c_{xx}[-m] = c_{xx}[m]$,

$$c_{xx}[m] = \begin{cases} \sum_{n=0}^{L-|m|-1} x[n]x[n+|m|] & m \leq L-1, \\ 0 & otherwise, \end{cases} \tag{2.8}$$

The implementation of this estimate could be implemented via using fast N-point DFT algorithm three times.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)kn} \tag{2.9}$$

$$|X[k]|^2 = X[k]X^*[k] \tag{2.10}$$

$$c_{xx}[m] = \frac{1}{N}\sum_{k=0}^{N-1} |X[k]|^2 e^{j(2\pi/N)km} \tag{2.11}$$

Finally, we could calculate the PSD of input signal via estimate of autocorrelation $\hat{\varphi}_{xx}[m]$.

$$s[m] = \begin{cases} \hat{\varphi}_{xx}[m]w_c[m] & 0 \leq m \leq Q-1 \\ 0 & Q \leq m \leq P-Q \\ \hat{\varphi}_{xx}[P-m]w_c[P-m] & 1 \leq P-m \leq Q-1 \end{cases} \tag{2.12}$$

The resulted PSD for given sliced signal is

$$S[k] = \sum_{m=0}^{Q-1} s[m]e^{-j(2\pi/Q)km} \tag{2.13}$$

Along with the moving of slicing window, we can acquire the spectrogram of input signal via STFT-CS.

We implement the algorithm with linear frequency modulation (LFM) to test it validity. The sample rate of the signal is 8000Hz, sweepping frequency from 0Hz to 4000Hz in ten seconds, which can be observed from the top row of Figure 2.1. The configurations of important variables corresponding to the Table 1 include: the length of slicing window is 256 sample points (sp), the length of STFT is 512 sp, the processed length of autocorrelation is 255 sp, the window function is Hamming window with window size equal to the size of Fourier transform.

As shown in the Figure 2.1, the first row is the standard algorithm to calculate STFT generating spectrogram, and the other rows are using the algorithm STFT-CS mentioned above. We can see that both standard STFT and STFT-CS could accurately trace the change of frequency.

Besides, based on the comparison of the lower three sub-figures in Figure 2.1 using SFTF-CS, we can see that as the increase of coprime pairs, there are more and more traces of aliasing frequency appearing in the spectrogram. This is because as the algorithm select less sample points to estimate the autocorrelation, there will be more variation.

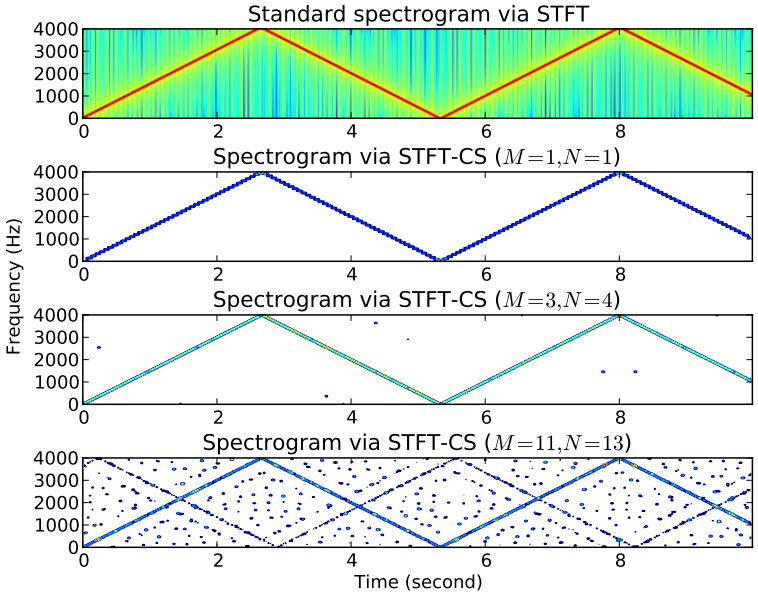

Figure 2.1: Comparison of dierent coprime pairs

On one hand, the decreasing of sample points is desirable for signal processing. For example, the fourth row in the Figure 2.1 only utilize about 17 percent of the sample points to achieve the same instantaneous PSD estimation with minor quality degeneration. On the other hand, however, the variation become more obvious if we continue increasing the values of coprime pair. This is the motivation for us to develop the 2-step STFT-CS presented in the next sub-section.

2.3.2 2-steps STFT coprime sampling

As the spectrogram described above, large values of coprime pair could generate lots of noise. An intuitive method to identify fundamental frequency buried under noise is to calculate its autocorrelation. Then, it becomes an interesting procedure of iterative autocorrelation, that is, estimating the autocorrelation via using convolution three times.

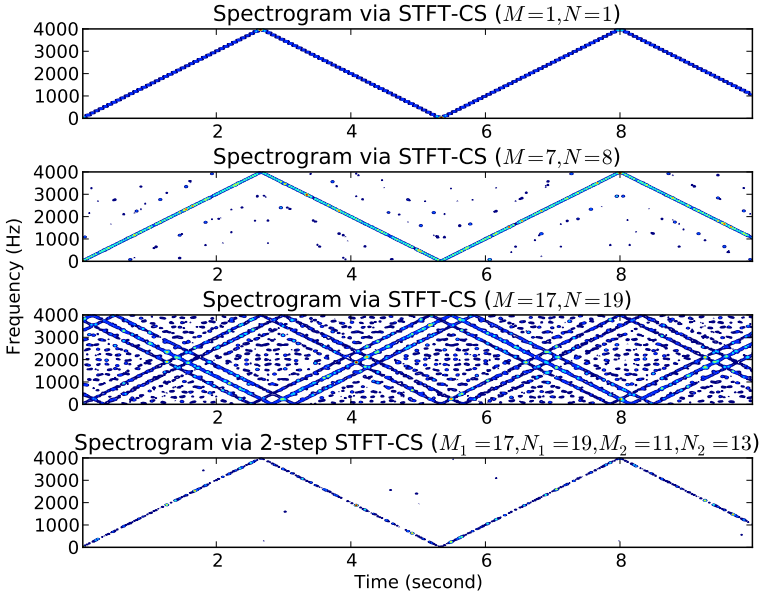

Figure 2.2: Comparison of STFT-CS and 2-step STFT-CS

In time domain, we calculate the autocorrelation function based on (2.7)

$$\hat{\varphi}'_{xx}[m] = \hat{\varphi}_{1xx}[m] * \hat{\varphi}^*_{2xx}[m] = \sum_{n=0}^{L-|m|-1} \hat{\varphi}_{1xx}[n]\hat{\varphi}^*_{2xx}[n+|m|] \tag{2.14}$$

where the $\varphi_{1xx}[m]$ and $\varphi_{2xx}[m]$ are two autocorrelation estimations which could be either same or different values of coprime pairs. The counterpart in frequency domain is straightforward. It is the product of PSD generated by two coprime pairs.

$$S'[k] = DFT(\hat{\varphi}_{1xx}[m] * \hat{\varphi}^*_{2xx}[m]) = S_1[k]S_2[k] \tag{2.15}$$

In Figure 2.2, we show the result of 2-steps STFT-CS together comparing with three results of 1-step STFT-CS with different configurations. The first row lists STFT-CS without coprime sampling as benchmark. The second and third rows are consistent with what we found in the previous sub-section. When the coprime pair increases to 17 and 19, we can hardly distinguish the real trace of spectrogram from the noise aliasing. The fourth row is the result of 2-steps STFT-CS using $M_1 = 17$,

$N_1 = 19$, and $M_2 = 11$, $N_2 = 13$. The resulting sequence has roughly the same degree of down-sampling rate (about 27 percent of the original sample points) as the experiment in second row. But we can observe that via the 2-step autocorrelation the false positive PSD estimates are obviously decreased.

2.3.3 Variation analysis for estimating autocorrelation

In the paper [27], the coprime sampling is the method dealing with the sub-Nyquist sampling frequency. Though it does provide promising potential of dramatically decreasing the sampling rate via coprime pair, the estimation is inherently suffering the problem of taking much longer latency. While in the non-stationary scenario, this situation would raise the major problem generating pronounced estimation variation for the reason that only a small piece of samples could be considered as stationary and processed once with autocorrelation estimation in STFT-CS. There is not enough latency permitted for averaging.

In other words, the statistical stability is sacrificed negatively proportional to the degree of coprime sampling. As the choice of coprime pair increases, the density of differential array generated would decrease correspondingly, though the coprime sampling might still calculate the full coverage of all lag by satisfying the *Property 4*. Then the correlation estimates at that lag could be deteriorated offsetting from the real values.

The paper [12] examined the error of estimating autocorrelation and the paper [13] linked the variation with sampling rate and refined it in the form of mean-square error. Besides, this paper also advocated that for short data records, whose sample points were less than 50 or the product of bandwidth and sampling period is less than 25, the preferred sampling rate was the twice of Nyquist rate. Otherwise, there would be obvious increases in the variance of the estimation.

Comparing this claim with the scenario of experiments in this paper, the sampling periods would fall into the category of short-data records while the sampling rate should be regarded as sub-Nyquist rate which is much lower than the desired rate in this criteria. Hence, the estimation will definitely suffer from significant variance.

The method of statistial differential could be used for estimating the covariances of autocorrelation coefficients [10]. For convenience of analysis, we could treat the LFM as piecewise stationary signal and define it as

$$x[n] = \sum_{s=0}^{L} h_s \varepsilon_{n-s} \tag{2.16}$$

where the series $\sum_{n=0}^{\infty} h_n$ are absolutely convergent, and ε_n is a WSS process with zero means and variance δ^2, that is

$$E(\varepsilon_n) = 0, \quad E(\varepsilon_n^2) = \delta^2 \tag{2.17}$$

$$E(\varepsilon_n \varepsilon_m) = 0, \quad for\ m \neq n \tag{2.18}$$

Then, the real value of autocorrelation is

$$R_k = cov(x_n, x_n + k) = \delta^2 \sum_{s=0}^{L} h_s h_{s+k} \tag{2.19}$$

and the estimation of autocorrelation is

$$\hat{R}_{k,L} = E(x_{n_1 M} x_{n_2 N}) \tag{2.20}$$

standing for averaging all of the available values of $x_{n_1 M} x_{n_2 N}$ to calculate the autocorrelation k within the range L.

Assume $h_t = 0$, we could calculate the covariance based on (2.16)

$$\begin{aligned} cov(x_n x_{m+k}, x_p x_{q+k}) =& \kappa_4 \sum_{r=-\infty}^{+\infty} h_{n-r} h_{m+k-r} h_{p-r} h_{q+k-r} \\ &+ R_{n-p} Rm - q + R_{n-q-k} R_{m-p+k} \end{aligned} \tag{2.21}$$

where $\kappa_4 = E(\varepsilon^4) - 3\delta^4$.

Therefore, we could have [11]

$$\lim_{L\to\infty} Lcov(\hat{R}_{k,L}; \hat{R}_{l,L}) = \frac{\kappa_4}{\delta^4} R_k R_l + \sum_{q=-\infty}^{+\infty} (R_q R_{q+k-l} + R_{q+k} R_{q-l}) = v_{kl} \tag{2.22}$$

and the particular case is the variance of autocorrelation

$$\lim_{L\to\infty} Lvar(\hat{R}_{k,L}) = \frac{\kappa_4}{\delta^4} R_k^2 + \sum_{q=-\infty}^{+\infty} (R_q^2 + R_{q+k} R_{q-k}) = v_{kk} \tag{2.23}$$

Another estimator for the autocorrelation is

$$C_{k.L} = \frac{1}{L-k} \sum_{l=1}^{L-k} x_t x_{t+k} \tag{2.24}$$

which confines estimate only based on the available sample points.

Similarly to (2.22, 2.23), we have

$$\lim_{L\to\infty} L\, cov(C_{k,L}; C_{l,L}) = v_{kl} \tag{2.25}$$

$$\lim_{L\to\infty} L\, var(C_{k,L}) = v_{kk} \tag{2.26}$$

Compared with (2.22), we could have

$$var(C_{k,L} - \hat{R}_{k,L}) = O(\frac{1}{L^2}) \tag{2.27}$$

Based on (2.23, 2.26) with Schwarz Inequality, we could have two measures for the variation of autocorrelation estimation with the lengh of available sample points.

$$cov(C_{k,L}; C_{l,L}) - cov(\hat{R}_{k,L}; \hat{R}_{l,L}) = O(\frac{1}{L^{3/2}}) \tag{2.28}$$

$$E(\hat{R}_{k,L} - R_k) = O(\sqrt{\frac{v_{kk}}{L}}) \tag{2.29}$$

From (2.28, 2.29) we can see why the estimate variation is increase as the decrease of sample points. This is an inherent problem confining the choices of coprime pairs in processing non-stationary signal using coprime sampling.

2.4 Implementation in radar signal processing

The working principle of matched filter in radar signal processing is to output the cross-correlation of target-plus-noise signal and transmitted signal [14]. So, it is possible to implement the matched filter as a correlation process. When the signal-to-noise (SNR) ratio is large, the output of the matched filter can usually be approximated be the autocorrelation function of the transmitted signal. Hence, we could use much less sampling points via coprime sampling to estimation the output of matched filter.

In this Section, we still consider the typical LFM waveform, which is consistent with the previous section and also used as a basic waveform in radar transmission because it could independently control pulse energy through its duration and range resolution through its bandwidth [15]. Thus, if the transmitted signal could be processed to have long duration and narrowly concentrated autocorrelation, both good range resolution and good energy can be obtained simultaneously.

Considering a modified waveform $x'(t)$ by modulating $x(t)$ with a LFM complex chirp and compute its complex ambiguity function

$$x'(t) = x(t)e^{j\pi\beta t^2/\tau} \tag{2.30}$$

The instantaneous frequency of this waveform is the derivative of the phase function

$$F_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = \frac{\beta}{\tau}t \tag{2.31}$$

in which the $\beta\tau$ is called time-bandwidth product of the LFM pulse. The time-delay measurement error is proportional to τ and the frequency measurement error is proportional to $1/\tau$.

In many radar application, the moving target generate Doppler shift in its echo signal, which makes the output of the matched filter should be considered as the cross

correlation between the Doppler-shifted received signal and the transmitted signal. In this case, we use ambiguity function (AF) to generate the behavior of a waveform paired with its matched filter. Based on the analysis of AF, we could easily examining resolution, side lobe behavior, and ambiguities in both time and Doppler domains.

Assume the Doppler frequency is F_D, then the input waveform with a Doppler-shifted response is $x(t)e^{j2\pi F_D t}$. Also assume that the filter is designed to peak at $T_M = 0$, which means that the time axis at the filter output is relative to the expected peak output time for the range of a target. Assuming M and N are the coprime pair and T_s is the sampling rate. Then the AF could be defined as

$$\begin{aligned}
\hat{A}'(k, F_D) &= \int_{-\infty}^{\infty} x'(MT_s)x'^{*}(NT_s)e^{j2\pi MT_s F_D} ds \\
&= \int_{-\infty}^{\infty} x(MT_s)x^{*}(NT_s)e^{j\pi\beta(M^2-N^2)T_s^2/\tau + j2\pi MT_s F_D} ds \\
&= e^{-j\pi\beta k^2/\tau} \int_{-\infty}^{\infty} x(MT_s)x^{*}(NT_s)e^{j2\pi(F_D+\beta k/\tau)MT_s} ds \\
&= e^{-j\pi\beta k^2/\tau} \hat{A}(k, F_D + \frac{\beta}{\tau}k)
\end{aligned} \tag{2.32}$$

where k is the difference between two sample points, and $\hat{A}(k, F_D)$ is the original complex ambiguity function for the simple pulse signal

$$\hat{A}(k, F_D) = \frac{e^{j2\pi F_D k/2}}{\tau j 2\pi F_D}(e^{j2\pi F_D \frac{\tau-k}{2}} - e^{-j2\pi F_D \frac{\tau-k}{2}}) \tag{2.33}$$

And its amplitude is

$$A(k, F_D) = \left|\hat{A}(k, F_D)\right| = \left|\frac{\sin\left[\pi F_D(\tau - |k|)\right]}{\tau\pi F_D}\right| \tag{2.34}$$

Then we can have the amplitude for the AF of the LFM waveform

$$A'(k, F_D) = |\hat{A}'(k, F_D)| = \left|\frac{\sin\left[\pi(F_D + \beta k/\tau)(\tau - |k|)\right]}{\tau\pi(F_D + \beta k/\tau)}\right| \tag{2.35}$$

The zero-Doppler cut of the LFM ambiguity function, which is just the matched filter output when there is no Doppler mismatch, is

$$A'(k,0) = \left| \frac{\sin\left[\pi\beta k(1-|k|/\tau)\right]}{\pi\beta k} \right| \quad -\tau \leq k \leq \tau \tag{2.36}$$

and the zero-delay response is

$$A'(0,F_D) = \left| \frac{\sin(\pi F_D \tau)}{\pi F_D \tau} \right| \tag{2.37}$$

In the experiment, we use coprime sampling on both transmitted signal in matched filter and received signal. Because the length of the chirp is predefined and need to fully analyze, based on *Property 2*, we could only have the difference co-array of index with missing values. But since the missing values will be more often for the autocorrelation with larger values, and we have already assumed $T_M = 0$ making the AF located relative to the time axis, there is not obvious effect of the missing values for the image generated by coprime sampled AF. The following simulation also confirms this claim.

From Figure 2.3, we can see that when we use small values of coprime pair in the upper right plotting, the resulting AF has inconspicuous degradation comparing with the upper left one, which is derived directly from formula. But as the values of coprime pair increase, there will be duplicated aliasing parts getting closer to the correct estimation. When we choose $M = 9$ and $N = 7$, the aliasing parts could still be easily eliminated, but when the pair becomes $M = 10$ and $N = 11$, or even bigger, the resulting AF is unable to use because all of the estimations overlap with each other.

Then, based on Figure 2.4 and Figure 2.5, we can observe different effects of the coprime sampling to the estimate of Doppler shift and time delay. Both of them are generated simutanously with Figure 2.3. In Figure 2.4, because the coprime sampling

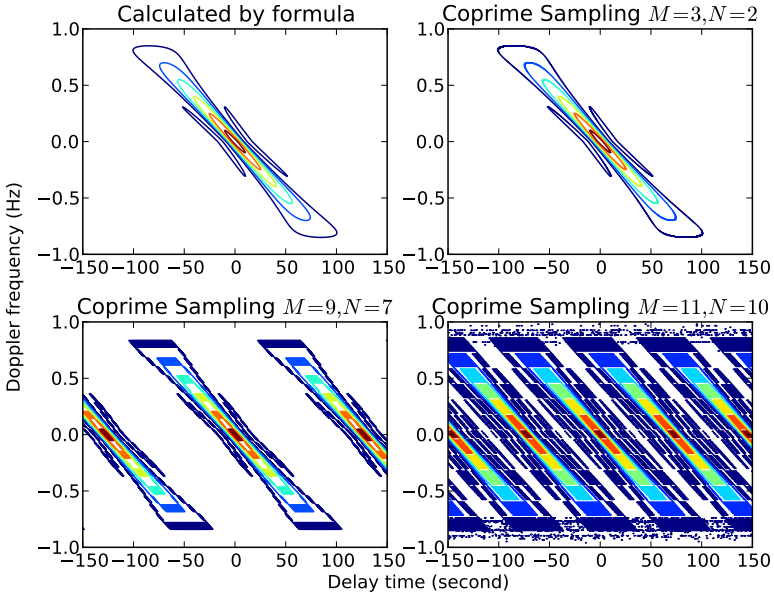

Figure 2.3: Comparison of dierent choices of coprime sampling for AF

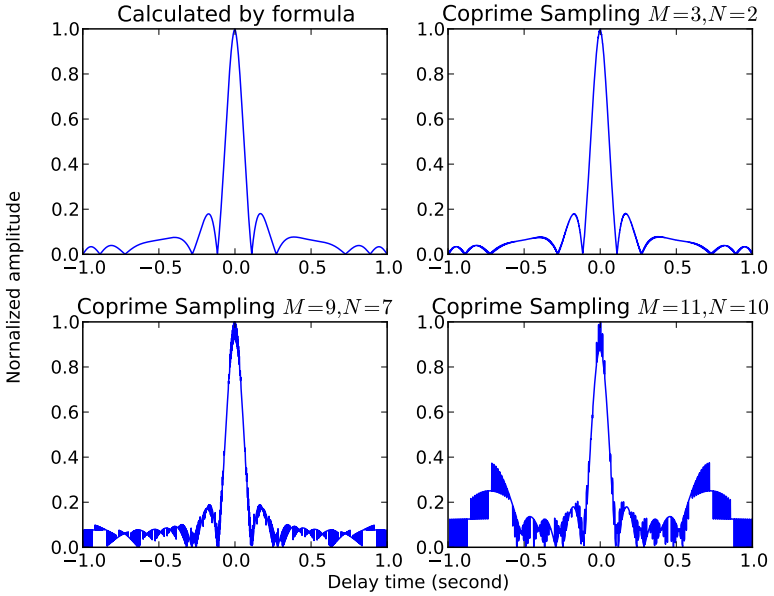

Figure 2.4: Zero Doppler shift Ambiguity Function

is implemented in the time domain, the variation becomes more and more obvious as the increase of coprime factors. We have thoroughly discuss the reason of this

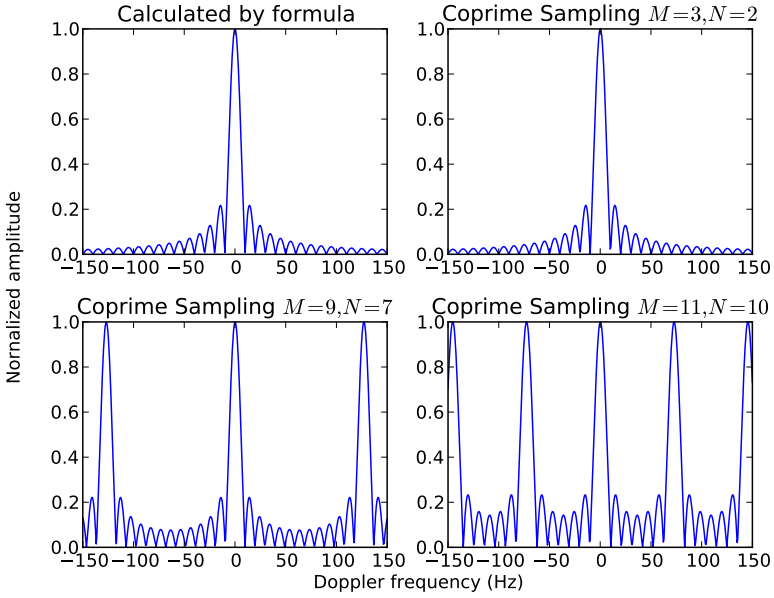

Figure 2.5: Zero-delay Ambiguity Function

phenomenon in the previous section. In Figure 2.5, since we keep the iteration along the Doppler axis the same, there is no variation existing. As the values of coprime pair increase, however, the distance between Doppler shift becomes smaller and smaller. Hence, we can conclude that as the increase of values of coprime pair, it will have deleterious effects including amplifying variation along time axis and decrease the scope of Doppler shift frequency.

To further quantify the effect of coprime sampling, we enumerate all coprime combinations under 17. The reason that we choose the threshold as 17 is because if the values of pair above this threshold servere overlaping of aliasing parts make the ouput useless. Besides, as shown in the following experiments, we find most of the results could be consistently arranged according to the products of coprime pairs. That is, four out of five important properties of coprime sampling AF are relevant with the product of coprime pairs rather than the value of either factor.

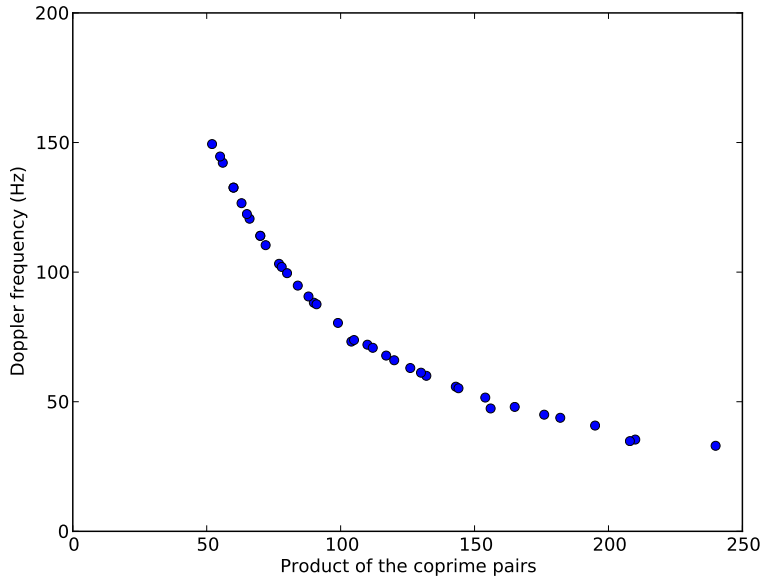

Figure 2.6: Distance between nearest main lobes in Doppler axis

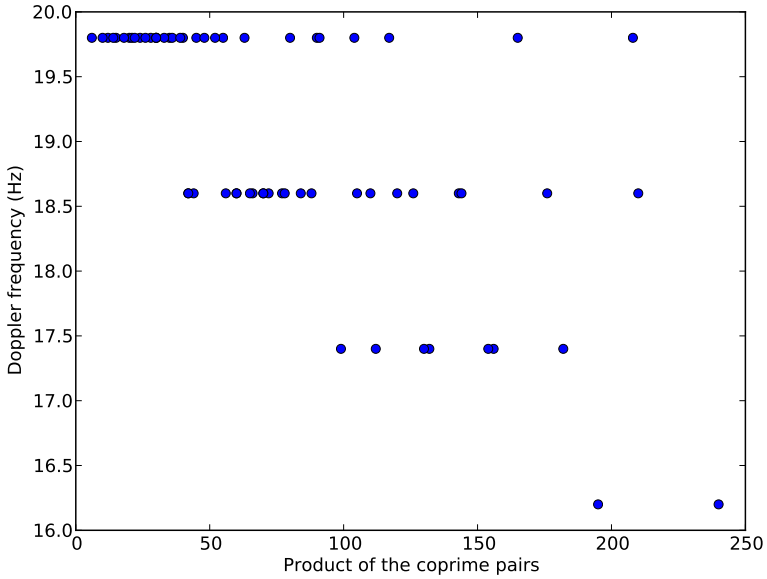

Figure 2.7: The width of major main lobe in Doppler axis

The distance between main lobes in Doppler axis determines the scope of Doppler frequency. From Figure 2.6, we can see that this distance is decreasing

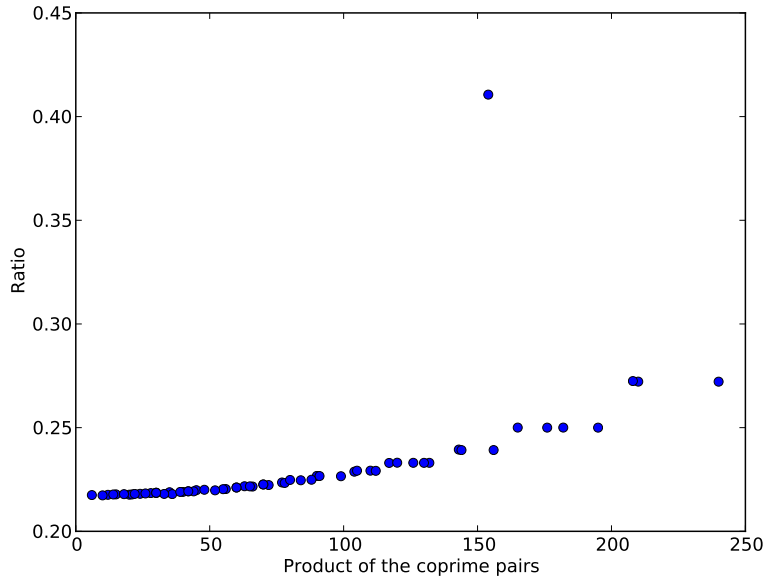

Figure 2.8: Ratio of 2nd side lobes to main lobes in Doppler axis

monotonically from out-of-scope to about 33Hz along with the increase of the product of coprime pair. Considering the width of main lobe provided in Figure 2.7, for the case of 33Hz distance, the second lobes of two AF estimations would overlap together. Note that for product less than 50, there will be no duplicated main lobe in the scope. For the worst case, the largest side lobes of each duplicate have overlapped together.

The width of the main lobe in Doppler axis determines the Doppler resolution. In the Figure 2.7, its range is from 19.8 Hz to 16.2 Hz. The width has only three discrete possible values and does not directly relevant with the product of coprime pair, though the general trend of width is getting smaller with larger products. This finding is instructive to find such coprime pair with narrow main lobe width but also less variation in time domain and longer distance among main lobes in Doppler axis.

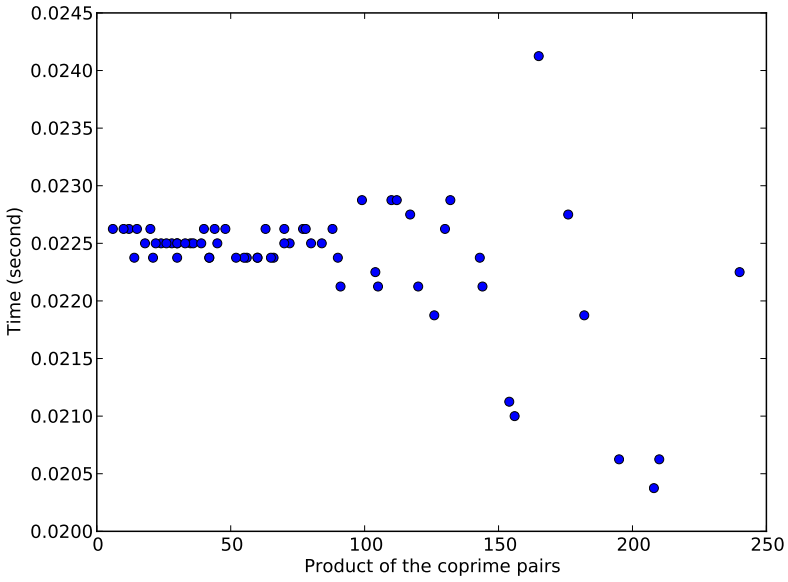

Figure 2.9: The width of the main lobe in time domain

Despite the largest side lobes in the Doppler axis become larger along with increasing coprime pairs, as shown in Figure 2.8, this is still not the major challenge comparing with the main lobes approaching to each other shown in Figure 2.6. Note that there is one abnormal value generated by M=14 and N=11. But it is more like a cutting-off main lobe located in the edge of scope rather than a real side lobe.

In Figure 2.9, the radiated shape shows no obvious relationship between the trend of main lobe and the choice of coprime pairs.

Comparing Figure 2.10 with Figure 2.8, we can see the main problem in time domain is caused by the variation, which in turn make the largest side lobes comparable to the main lobe. Note that there is a turning point in the production of 88 for the ratio changing from stable around 18 percent to increasing with the production.

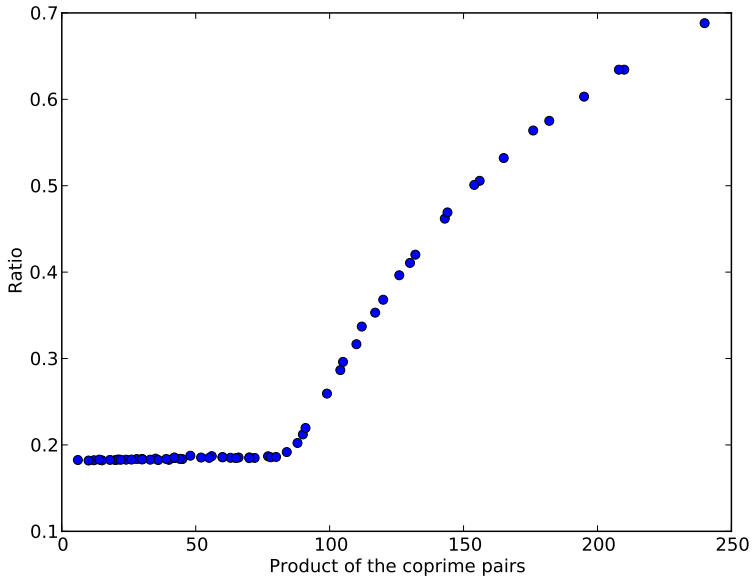

Figure 2.10: Ratio of 2nd side lobes to main lobes in time domain

2.5 Conclusions

In this Chapter, we develop the algorithm STFT-CS to deal with non-stationary signal. The decreasing of processed data is favorable for sparse sampling as well as decreasing the computation complexity, but the cost is increasing estimate variation. To alleviate the side-effects, we introduce two-steps STFT-CS. The simulation indicates it is effective to eliminate aliasing estimations.

Besides, we also implement the coprime sampling with the matched filter of radar signal processing, and quantify the effect of coprime sampling in such process. Based on our analysis, one could integrate the coprime sampling in radar system to detect targets, and choose the suitable configuration based on specific circumstance and needs.

The future research directions include further optimizing the algorithm and using it with real-world radar data. Besides, coprime sampling and coprime sensor array

do have many interesting features which might be useful for other applications, such as wireless communication or image/audio signal processing. Moreover, just as using STFY-CS converting time domain signal to more meaningful PSD representation, coprime sampling could be regarded as preprocessing for contaminant data to restore the fundamental information.

CHAPTER 3

Coprime sampling of HOS for channel order determination

3.1 Introduction

The ultra-wideband (UWB) radio receives lots of attention in wireless communication, because it possesses many properties of a spread-spectrum radio link, such as multiple access addressing, interference suppression, and low probability of detection and interception. Its indoor channel can be approximated with sufficient accuracy using an autoregressive moving-average (ARMA) model [36].

To estimate the parameters of an ARMA model, higher-order statistics (HOS) are preferable to second-order statistics, because the former approaches reveal both amplitude and phase information, and are blind to any kind of Gaussian processes [23]. Hence, HOS are able to extract rich information from the unknown model and meanwhile boost SNR by disregarding both white and color Gaussian measurement noise [37]. Among works that utilize HOS in parameter estimation, the order of a model is often assumed to be known beforehand. But in most realistic situations, the exact order of the model is not known and must be estimated prior to solving the parameter estimation problem. The order determination is an indispensable part of any parametric modeling procedure.

In general, existing order determination methods are autocorrelation-based [38, 39] or HOS-based [40–42]. They can also be divided by inherent theories–using information theoretical criteria [39, 40] or linear algebraic methods [38].

Based on second-order statistics, Tugnait [39] proposed a heuristic modification of Akaike's information criterion (AIC) for order selection. It was useful for the mini-

mum phase systems or selecting the spectrally equivalent minimum phase counterpart to the given possibly nonminimum phase system. Similar algorithms were developed for estimating the orders of two-dimensional autoregressive (AR) and ARMA models [38], as well as estimating the 3-D AR model with rank test procedure in singular value decomposition (SVD) [43]. The paper [44] further derived factorization of covariance matrix and recursive estimation for multidimensional models with different order in each dimension. However, this kind of approaches produced biased results when additive Gaussian noise was present [45]. Besides, they assumed Gaussian process and/or minimum phase models, and could not estimate all-pass factors inherent in ARMA models.

Based on asymptotic property of third-order cumulants, the paper [46] proposed an algorithm conducting model validation and order selection at the same time. The AR order could also be determined using cumulants with the minimum description length (MDL) criterion [47]. For 2-D models, the order of the 2-D moving-average (MA) systems were estimated iteratively by minimizing HOS cost functions [42], and the 2-D AR models were estimated by solving Hankel matrix constructed by HOS cumulants [45]. Besides, Giannakis [48] developed two methods for non-Gaussian ARMA processes, one of which used Gram-Schmidt procedure performing a linear dependency search among the columns. Another method achieved better noise robustness using SVD to select AR order and polyspectral phases to estimate the relative MA degree. Zhang [41] improved the SVD performance in low SNR using the product of diagonal entries test. In addition, information theoretic criteria were derived for the kth-order cumulant [40], which did not require subjective thresholds and yielded consistent estimates for both the orders and the parameters.

Given that the order determination is one important yet primitive stage to process the signal, the motivation of Chapter 3 is to use co-prime sampling decreasing

the computational overhead while still retaining all merits of cumulant-based order selection, including determining the order of ARMA model with phase-only factors and resisting Gaussian noise with unknown variance. Furthermore, as indicated by previous research about co-prime sampling [25], the co-prime sampled sequences generate larger variance, which makes it more susceptible to noise and order mismatch. This characteristic provides another insight to double check the order estimates.

In Section 3.2, we discuss the theoretical background of both co-prime sampling and HOS. In Section 3.3, we develop theorems of pairwise co-prime sequences (PCS) for estimating HOS. We describe the algorithm of order determination using PCS-based HOS in Section 3.4, and provide simulation results in Section 3.5. Finally, we draw the conclusions in Section 3.6.

3.2 Preliminaries

In this section, we provide an overview of higher-order statistics and co-prime sampling, and demonstrate their relationship based on the estimation of second-order moment.

3.2.1 Definition of higher-order statistics

Assuming $x(t)$ is a zero-mean random process with at least kth-order stationary, the kth-order cumulant of this process is defined as the first $k-1$ coefficients of Taylor

series expansion of the data set. The expectations of second-order, third-order, and fourth-order cumulant are [23]

$$\begin{aligned} c_{2x}(\tau) =& E[x(t)x(t+\tau)], && (3.1)\\ c_{3x}(\tau_1,\tau_2) =& E[x(t)x(t+\tau_1)x(t+\tau_2)], && (3.2)\\ c_{4x}(\tau_1,\tau_2,\tau_3) =& E[x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)] \\ & - c_{2x}(\tau_1)c_{2,x}(\tau_2-\tau_3) \\ & - c_{2x}(\tau_2)c_{2,x}(\tau_3-\tau_1) \\ & - c_{2x}(\tau_3)c_{2,x}(\tau_1-\tau_2). && (3.3) \end{aligned}$$

The *kth*-order spectrum is defined as the Fourier transform of its *kth*-order cumulant:

$$S_{kx}(\omega_1,\omega_2,\ldots,\omega_{k-1}) = \sum_{\tau_1=-\infty}^{+\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{+\infty} c_{kx}(\tau_1,\tau_2,\ldots,\tau_{k-1}) \cdot \exp\left(-j\sum_{i=1}^{k-1}\omega_i\tau_i\right). \tag{3.4}$$

Specifically, the third-order spectrum is also called bispectrum, and the fourth-order spectrum is trispectrum.

3.2.2 Properties of coprime sampling

Co-prime sampling can be imagined as a pair of uniform samplers simultaneously receiving a discrete-time signal $w(n)$. The rates of two samplers are a co-prime pair M and N.

$$\begin{aligned} w_1(a_1) &= w(Ma_1) \\ w_2(a_2) &= w(Na_2). \end{aligned} \tag{3.5}$$

Except for the indexes equal to the least common multiples, these two sequences do not have any overlapped mapping to the input signal $w(n)$, but the difference of the mapped indexes $Ma_1 - Na_2$ have seamless coverage in $[0, MN]$. This difference

set naturally relates the co-prime sampled sequences $w_1(a_1)$ and $w_2(a_2)$ to estimating the second-order moments:

$$\hat{c}_{2w}(k) = \frac{1}{L}\sum_{l=0}^{L-1} w_1(M(a_1+Nl))w_2^*(N(a_2+Ml)), \tag{3.6}$$

where $k = Ma_1 - Na_2$.

It is worth noting that there is no constraint on the choices of downsampling rate M and N. Theoretically, sample rates for the intermediate two sequences are not restricted by the Nyquist rate and the sample points can be arbitrarily sparse. Besides, the ranges of M and N are also essential for the statistical characteristics of the derived autocorrelation. In particular, co-prime sampling has two basic properties:

Property 4. *When $0 \le a_1 \le N-1$ and $0 \le a_2 \le M-1$, the values of $k = Ma_1 - Na_2$ are restricted to the range $-(M-1)N \le k \le (N-1)M$. Since there are at most MN distinct values for k, the co-prime sampled sequences cannot estimate all autocorrelation coefficients in the given range.*

Property 5. *Given sample points in the range $(-L, L)$, the largest co-prime pair M and N subjected to $MN < L$, and a_1 and a_2 in the range $0 \le a_1 \le \lfloor L/M \rfloor$ and $-\lfloor L/N \rfloor \le a_2 \le \lfloor L/N \rfloor$, the resulting difference indexes $k = Ma_1 - Na_2$ are able to achieve full coverage in the range $0 \le k \le L-1$. Here $\lfloor X \rfloor$ stand for the largest integer no larger than X.*

The proofs can be referred in [25, 27].

3.3 Pairwise co-prime sequence

3.3.1 Sufficient condition for HOS derived from PCS

In order to implement co-prime sampling with HOS, we extend the co-prime sampling pair to pairwise co-prime sequences (PCS). In this Section, we formally prove

two theorems of PCS. First of all, it is sufficient for any index value k of rth-order statistics calculated from PCS.

Theorem 1. *Given pairwise co-prime integers $n_1, n_2, \ldots, n_r$ ($gcd(n_i, n_j) = 1$ for all $i \neq j$) and any integer k, we can always find integers $a_1, a_2, \ldots, a_r$ such that*

$$k = \sum_{i=1}^{r} a_i n_i. \tag{3.7}$$

Proof. Let us first check the existence of $r = 2$ and $k = 1$, that is $a_1 n_1 + a_2 n_2 = 1$. Based on the Euclidean Algorithm [30], we derive the value of a_1 and a_2

$$\begin{aligned} gcd(n_1, n_2) = r_n =& r_{n-2} - r_{n-1} q_n \\ =& r_{n-1}(1 + q_n q_{n-1}) - r_{n-3} q_n \\ =& r_{n-3} \cdot * + r_{n-4} \cdot * \\ & \vdots \\ =& n_2 \cdot * + r_1 \cdot * \\ =& n_1 \cdot * + n_2 \cdot * \end{aligned} \tag{3.8}$$

where $*$ stands for integers derived from the previous step, and r_i stands for the remainder in the next iteration. Hence, we can always find a_1 and a_2, satisfying $a_1 n_1 + a_2 n_2 = 1$. For any value of k_2, a_1' and a_2' also exist satisfying $a_1' n_1 + a_2' n_2 = k_2$.

Hence, for any $k_3 = a_3 n_3 + k_2 = a_3 n_3 + a_1 n_1 + a_2 n_2$, we can always find the means to combine three prime numbers n_1, n_2, and n_3. Suppose it holds through 1 to r'. That is $k_{r'} = a_1 n_1 + a_2 n_2 + \cdots + a_{r'} n_{r'}$. Similarly, for $r = r' + 1$, $k_r = a_{r'+1} n_{r'+1} + k_{r'} = a_1 n_1 + a_2 n_2 + \cdots + a_r n_r$. Hence, it holds for all $k, a_1, a_2, \ldots, a_r \in \mathbb{Z}$. □

3.3.2 Full coverage of HOS indices based on PCS

The following theorem proves that certain PCS scheme achieves seamless HOS coverage for a finite segmentation.

Theorem 2. *Given pairwise co-prime integers $n_1, n_2, \ldots, n_r$ ($gcd(n_i, n_j) = 1$ for all $i \neq j$), let $N = n_1 n_2 \ldots n_r$, and the index range be $a_i \in \left(-\frac{N}{n_i}, \frac{N}{n_i}\right)$ for $i = 1, 2, \ldots, r$ and $r \geq 2$. We can find solutions of indexes of $k = \sum_{i=1}^{r} a_i n_i$ for any integer k in the range $k \in [0, N-1]$.*

Proof. Based on Theorem 1, there exist coefficients b_i for $i = 1, 2, \ldots, r$, such that the index k of rth-order statistics can be represented as $k = \sum_{i=1}^{r} b_i n_i$.Then, the coefficients of the pairwise co-prime set is rewritten as

$$k = \sum_{i=1}^{r} (b_i - p_i \frac{N}{n_i}) n_i. \tag{3.9}$$

Considering the coefficient of n_1, if p_1 is properly chosen, because the step is $\frac{N}{n_1}$, the result coefficient $a_1 = b_1 - p_1 \frac{N}{n_1}$ is in the range $\left(-\frac{N}{n_1}, \frac{N}{n_1}\right)$.

For the case that $r = 2$, let $a_2 = b_2 - p_2 \frac{N}{n_2}$, and $a_2 n_2 = k - a_1 n_1$. Because $k \in [0, N-1]$ and $a_1 n_1 \in (-N, N)$, $a_2 \in \left(-\frac{N}{n_2}, \frac{N}{n_2}\right)$. Similarly, we could deduce that $a_l \in \left(-\frac{N}{n_l}, \frac{N}{n_l}\right)$ for $l \geq 2$. The range also holds for $a_{l+1} \in \left(-\frac{N}{n_{l+1}}, \frac{N}{n_{l+1}}\right)$, which concludes the proof. □

3.3.3 Unique mapping from PCS to HOS indices

In addition, based on the simultaneous congruences in Chinese Remainder Theorem and Theorem 2, the following corollary guarantees the index in HOS can be uniquely calculated from a given set of PCS.

Corollary 1. (Pairwise co-prime mapping) *Given pairwise relatively prime integers $n_1, n_2, \ldots, n_r$ and their product $N = n_1 n_2 \ldots n_r$, the indexes a_i in the PCS sequences $n_i[a_i]$ can be regarded as simultaneous congruences from the index of higher-order statistics k to n_i, respectively.*

$$k \equiv a_i \pmod{n_i} \tag{3.10}$$

for $i = 1, 2, \ldots, r$. The solution for the set of congruences is unique modulo N, which is

$$k \equiv \sum_{i=1}^{r} a_i p_i \frac{N}{n_i} \pmod{N} \tag{3.11}$$

where the p_i are determined from

$$p_i \frac{N}{n_i} \equiv 1 \pmod{n_i}. \tag{3.12}$$

Proof. Denote $N_i = \frac{N}{n_i}$, for all i. Because $n_1, n_2, \ldots, n_r$ are PCS, $gcd(N_i, n_i) = 1$. Based on Theorem 1 and 2, we have

$$pN_i + qn_i = 1 \tag{3.13}$$

$$pN_i \equiv 1 \pmod{n_i} \tag{3.14}$$

for $0 \leq p < n_i$.

If $x \neq y$ and they satisfy the set of congruences, then

$$x - y \equiv 0 \pmod{n_i} \quad i = 1, \ldots, r. \tag{3.15}$$

Because $n_1, n_2, \ldots, n_r$ are pairwise co-prime, $x - y$ is also divisible by N. Therefore, x and y are identical modulo N. □

3.4 PCS-based HOS algorithm for ARMA order determination

3.4.1 Probelm formulation

Let us consider the causal linear signal model:

$$x(t) + \sum_{i=1}^{p} a(i)x(k-i) = \sum_{i=0}^{q} b(i)w(t-i) \tag{3.16}$$

where $\{a(i)\}_{i=1}^{p}$ and $\{b(i)\}_{i=0}^{q}$ denote the ARMA parameters, which can be further abbreviated as ARMA(p, q). The p and q are the order of AR and MA respectively. The observations consist of the signal $x(t)$ and the measurement noise $v(t)$:

$$y(t) = x(t) + v(t). \tag{3.17}$$

The transfer function in the Z domain is assumed to be free of pole-zero cancellations and exponentially stable:

$$H(z) = \sum_{i=0}^{+\infty} h(i) z^{-i} = \left[\sum_{i=0}^{q} b(i) z^{-i} \right] \Big/ \left[1 + \sum_{i=1}^{p} a(i) z^{-i} \right]$$
$$= B(z)/A(z) \tag{3.18}$$

Besides, we have three assumptions in the following content:

S.1 The driving noise $w(t)$ is non-Gaussian, zero mean, and independent and identical distributed (i.i.d.). Besides, it only has finite moments, which are estimated via expectations, and denoted as $E[w^2(t)] = \sigma_w^2$, $E[w^3(t)] = \gamma_{kw}$, and $E[w^6(t)] < +\infty$.

S.2 The measurement noise $v(t)$ is assumed to be zero mean, and has either white or color Gaussian distribution. It is also i.i.d. and independent of $w(t)$ with $E[v^2(t)] = \sigma_v^2$.

S.3 The upper bounds of p and q are known.

Given PCS $x[i]$, we are able to find any $\tau \in [0, N-1]$ based on Theorems 1 and 2, such that

$$\tau = \sum_{i=1}^{r} a_i n_i, \quad a_i \in \left(-\frac{N}{n_i}, \frac{N}{n_i} \right). \tag{3.19}$$

The second-order autocorrelation ($r = 2$) in (3.6) is rewritten as

$$\hat{c}_{2x}(\tau) = \frac{1}{L} \sum_{l=0}^{L} x_1(a_1 + \frac{N}{n_1} l) x_2(a_2 + \frac{N}{n_2} l) \tag{3.20}$$

where L denotes the number of segmentations used for averaging. Because the resulting indexes of PCS are calculated using combinations of different co-prime pairs rather than permutation of composite components, the estimates for theoretically symmetric points are not equal. So, they are further averaged within the segment via the symmetry of autocorrelation

$$c_{2x}(\tau) = \frac{1}{2} \left(\hat{c}_{2x}(\tau) + \hat{c}_{2x}(-\tau) \right). \tag{3.21}$$

The third-order cumulant ($r = 3$) is estimated as

$$\hat{c}_{3x}(\tau_1, \tau_2) = \frac{1}{L} \sum_{l=0}^{L} \prod_{i=1}^{3} x_i(a_i + \frac{N}{n_i} l). \tag{3.22}$$

Based on its symmetry property [23], the estimates are further averaged as

$$\begin{aligned} \hat{c}_{3x}(\tau_1, \tau_2) =& \frac{1}{6} \left[\hat{c}_{3x}(\tau_1, \tau_2) + \hat{c}_{3x}(\tau_2, \tau_1) + \hat{c}_{3x}(-\tau_2, \tau_1 - \tau_2) + \hat{c}_{3x}(-\tau_1, \tau_2 - \tau_1) \right. \\ & \left. + \hat{c}_{3x}(\tau_2 - \tau_1, -\tau_1) + \hat{c}_{3x}(\tau_1 - \tau_2, -\tau_2) \right]. \end{aligned} \tag{3.23}$$

3.4.2 AR order determination based on Hankel matrix

we use the theorem derived from Hankel matrix [48] and the third-order cumulant to select the AR order.

Theorem 3. *Define the* $(M_2 + 1)(N_2 - N_1 + 1) \times (M_2 + 1)$ *matrix of third-order cumulant* $\boldsymbol{C_e}$

$$\boldsymbol{C_e} = \begin{bmatrix} c_{3y}(M_1, N_1) & \dots & c_{3y}(M_1 + M_2, N_1) \\ \dots & \dots & \dots \\ c_{3y}(M_1, N_2) & \dots & c_{3y}(M_1 + M_2, N_2) \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ c_{3y}(M_1 + M_2, N_1) & \dots & c_{3y}(M_1 + 2M_2, N_1) \\ \dots & \dots & \dots \\ c_{3y}(M_1 + M_2, N_2) & \dots & c_{3y}(M_1 + 2M_2, N_2) \end{bmatrix}, \tag{3.24}$$

where

$$M_1 \geq q - p + 1$$

$$M_2 \geq p - 1$$

$$N_1 \leq q - p$$

$$N_2 \geq q.$$

The matrix $\boldsymbol{C_e}$ *has rank* p *if and only if the* $ARMA(p, q)$ *model is free of pole-zero cancellations.*

Because the singular values of a matrix are noise robust for rank estimation, SVD approach has been successfully applied to AR order determination. Theoretically, using SVD with matrix in Theorem 3 is to find the nonzero singular values in the matrix $\boldsymbol{C_e}$, but all the singular values are nonzero because of the estimation error and existence of noise. Consequently, the subjective rule to estimate p is to find the smallest difference between consecutive singular values $\{s_{p+1}^{\boldsymbol{C_e}} - s_{p+2}^{\boldsymbol{C_e}}\}$.

3.4.3 MA order determination based on HOS properties

To estimate the order of MA models, rather than using the algorithms mentioned in Section 3.1, we apply a tractable approach based on the basic statistical characteristic of HOS cumulant. After removing the AR impact from an ARMA sequence, the residual sequence can be regarded as a MA process. Suppose its order is q, then it is safe to infer that

$$\hat{c}_{3y}(q, 0) \neq 0, \quad \hat{c}_{3y}(q + 1, 0) \approx 0. \tag{3.25}$$

That is, the value of q is identified as the lag of the last cumulant sample that is not approximately equal to zero. Note that the performance of this method is directly related with the asymptotic property of the third-order cumulants derived from the

sampled sequence. The asymptotic covariance expressions in terms of the ARMA parameters can be found in [49]. However, in practice, we only estimate the sampled averages rather than their true values. Based on the assumption S.2, we have a consistent sample estimator of the covariance

$$\begin{aligned} cov[\hat{c}_{3y}(m,n),\hat{c}_{3y}(k,l)] \approx & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=-q+n}^{q+m}\left(1-\frac{|j|}{N}\right) \\ & \cdot[y(i)y(i+m)y(i+n)-\hat{c}_{3y}(m,n)] \\ & \cdot[y(i+j)y(i+j+k)y(i+j+l)-\hat{c}_{3y}(k,l)]. \end{aligned} \tag{3.26}$$

Let $k=m$ and $k=n$, we have the variance of sampled third-order cumulants

$$\begin{aligned} \hat{\sigma}^2[\hat{c}_{3y}(m,n)] \approx & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=-q+n}^{q+m}\left(1-\frac{|j|}{N}\right) \\ & \cdot[y(i)y(i+m)y(i+n)-\hat{c}_{3y}(m,n)] \\ & \cdot[y(i+j)y(i+j+m)y(i+j+n)-\hat{c}_{3y}(m,n)]. \end{aligned} \tag{3.27}$$

Because the impulse response of a stable ARMA model is exponentially decaying, its output cumulants can also be regarded as an MA process [41]. Hence, it is reasonable to assume $\hat{c}_{3y}(m,n)\approx 0$ for $m>q$ or $n>q$. The truncation point can be found by visual inspection or hypothetical testing.

3.5 Simulation results

3.5.1 Basic setting

In simulations, the driven noise $w(t)$ is a zero-mean exponential random deviate process with $E[w^2(t)]=1$ and $E[w^3(t)]=2$. We generate signal sequence with 5000 sample points for a single run, and average results from 50 times of Monte Carlo simulations for every scenario. Besides, we set the size of processing window is for 256 sample points and no overlap for consecutive windows. The estimates of each segment

Table 3.1: Order determination of a MA(3) model

stat.	$b(0)$	$b(1)$	$b(2)$	$b(3)$	$b(4)$	$b(5)$
Noise free environment (SNR=$+\infty$)						
c_{2y}	2.529	0.936	-0.314	-0.767	-0.006	-0.002
	± 0.011	± 0.003	± 0.002	± 0.002	± 0.002	± 0.003
c_{3y}	2.589	2.317	1.369	-1.555	-0.017	0.018
	± 0.220	± 0.108	± 0.032	± 0.040	± 0.014	± 0.013
c'_{3y}	2.672	2.352	1.389	-1.576	-0.039	0.026
	± 1.088	± 0.134	± 0.122	± 0.132	± 0.033	± 0.006
White Gaussian noise (SNR=0)						
c_{2y}	4.799	0.941	-0.321	-0.756	-0.016	-0.015
	± 0.023	± 0.008	± 0.004	± 0.004	± 0.005	± 0.007
c_{3y}	2.684	2.345	1.353	-1.606	-0.138	-0.048
	± 0.480	± 0.182	± 0.083	± 0.107	± 0.053	± 0.036
c'_{3y}	2.409	2.365	1.372	-1.699	-0.112	-0.074
	± 1.936	± 0.189	± 0.179	± 0.154	± 0.121	± 0.034
Colored Gaussian noise (SNR=0)						
c_{2y}	26.819	-9.336	$-2,791$	-3.939	9.110	-3.307
	± 0.817	± 0.428	± 0.464	± 0.296	± 0.705	± 0.408
c_{3y}	3.918	2.113	-1.114	-2.260	0.429	-0.198
	± 7.168	± 3.252	± 1.100	± 1.533	± 2.479	± 1.989
c'_{3y}	3.505	2.594	1.536	-2.256	-0.691	0.451
	± 7.160	± 1.748	± 1.711	± 2.601	± 2.540	$\pm 2,075$

are averaged to obtain the final result. In both test cases, we simulate scenarios including noise-free environment ($SNR = +\infty$), and white and color Gaussian noise at SNR levels from 1 ($0dB$) to 100 ($20dB$). Moreover, we select $n_1 = 1$, $n_2 = 2$, and $n_3 = 3$ in PCS, which decrease the complexity to less than 17% of the existing HOS algorithm.

3.5.2 MA model order determination

We simulate a MA(3) nonminimum phase model:

$$x(t) = w(t) + 0.9w(t-1) + 0.385w(t-2) - 0.771w(t-3) \tag{3.28}$$

which has zeros at $-0.75 \pm 0.85j$ and 0.6. The statistics of simulation are shown in Table 3.1. c_{2y} denotes the autocorrelation coefficients. c_{3y} denotes the cumulants calculated using existing HOS, and c'_{3y} are estimated via PCS-based HOS. The corresponding variances are listed below the averages.

The first part of Table 3.1 indicates all of the three algorithms perform well in the noise-free environment–the amplitudes of the fourth lags decrease dramatically compared with the third lags. The upper figure in Figure 3.1 visualizes these statistics and groups them according to different lags. Signals in the second part of Table 3.1 are corrupted by i.i.d. white Gaussian noise. Although the distinction between the third and fourth lags is still visible for the autocorrelation, its zero-lag autocorrelation is heavily affected. In contrast, the third-order cumulants reveal their insensitivity of white Gaussian noise. The lower histogram in Figure 3.1 shows the superiorities of HOS over autocorrelation in the environment with $0dB$ white Gaussian noise. From Figure 3.1, we also observe that the PCS-based HOS has negligible performance difference from the existing HOS algorithm. In the present of white Gaussian noise, both of them are able to derive the right order. But PCS-based algorithm is much simpler in terms of computational complexity.

Figure 3.2 visualizes the third part of the Table 3.1, which is the results simulated in $0dB$ color Gaussian noise. The noise is acquired by passing white Gaussian noise through MA channel with coefficients $[1, -2.33, 0.75, 0.5, 0.3, -1.4]$. The left vertical axis of Figure 3.2 is the amplitudes of cumulants, and its right axis is the ratio between variances and amplitudes linked as curves. Although the c_{2y} retains low variance, it is useless because the amplitudes of the fourth and fifth lags are large enough to conclude wrong answers. We can still use the amplitude of c_{3y} and c'_{3y} to derive the right order, but the curves of ratios provides better indicators, which

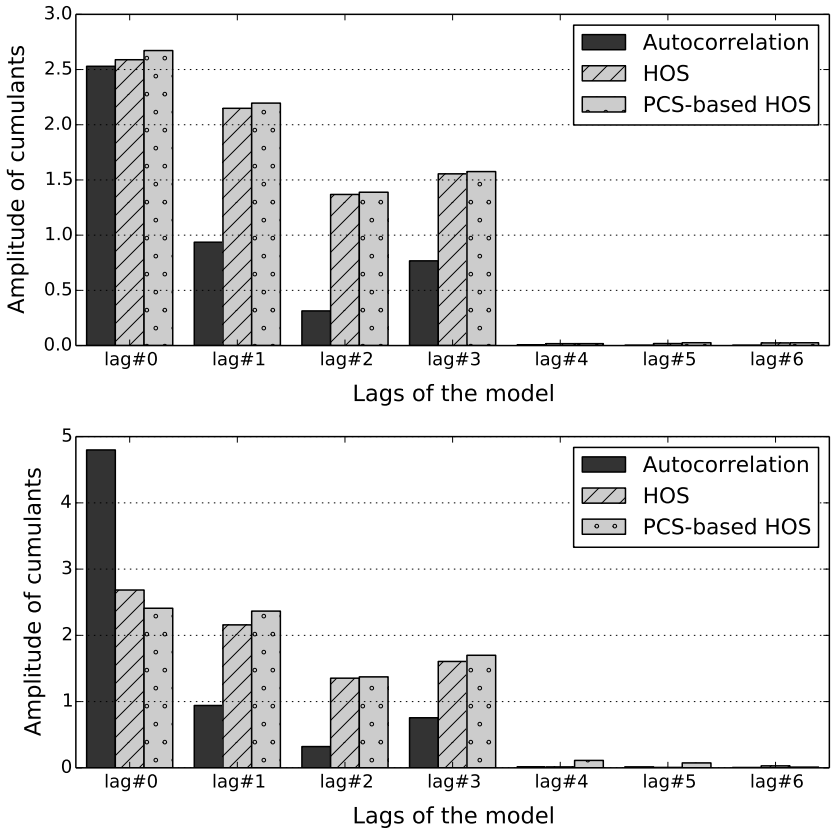

Figure 3.1: noise free environment

increase almost six times from the third to the fourth lag. The PCS-based HOS performs better for all mismatched orders.

3.5.3 ARMA model order determination

We simulate an ARMA(3,3) model with all-pass factors

$$\begin{aligned} & x(t) - 2.2x(t-1) + 1.77x(t-2) - 0.52x(t-3) \\ & = w(t) + 0.9w(t-1) + 0.385w(t-2) - 0.771w(t-3) \end{aligned} \tag{3.29}$$

with poles at $0.7 \pm j0.4$ and zeros at $-0.75 \pm j0.85$, 0.6. The noise $v(t)$ is the same as the previous case. The Table 3.2 is divided according to SNR levels. $S.V._{\boldsymbol{C_e}}$ stands for the normalized singular values derived from the matrix $\boldsymbol{C_e}$, and $S.V.'_{pcs}$ are derived

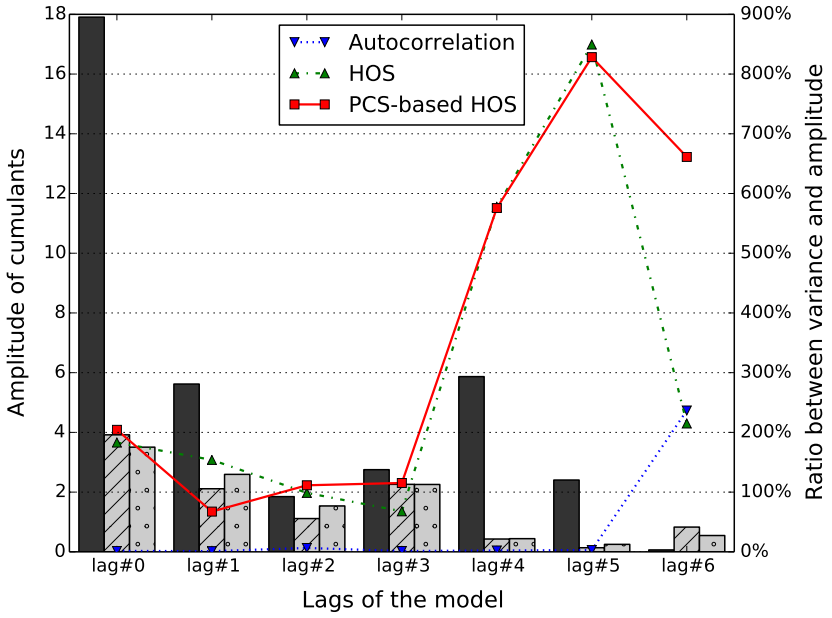

Figure 3.2: Colored Gaussian noise

from the same matrix but formed based on PCS-based HOS. c_{3y} and c'_{3y} are the same meaning as they are in Table 3.1. The columns for cumulants are lags of MA order started from zero, and the columns for singular values stand for their sorted order from 1 to 5.

From Table 3.2, we can see that both algorithms perform well for ARMA order determination in noise-free ($SNR = +\infty$) and mild noise ($SNR = 10$) environments. For the AR models, PCS has less effect after the SVD approach. The fourth and fifth singular values have the smallest difference. This indicates the right AR order is 3. For the MA order selection, PCS-based HOS produce larger variance, but the most obvious variances take place in the zeroth lag, which does not affect the final decisions of MA order.

Table 3.2: Order determination of a ARMA(3,3) model

Noise free environment (SNR=$+\infty$)					
$S.V._{\boldsymbol{c_e}}$	1.000	0.435	0.154	0.030	0.008
	$\pm$0.000	$\pm$0.014	$\pm$0.002	$\pm$0.000	$\pm$0.000
$S.V._{pcs}'$	1.000	0.465	0.265	0.103	0.078
	$\pm$0.000	$\pm$0.006	$\pm$0.004	$\pm$0.002	$\pm$0.001
c_{3y}	2.588	2.317	1.369	-1.555	-0.017
	$\pm$0.220	$\pm$0.108	$\pm$0.032	$\pm$0.040	$\pm$0.014
c_{3y}'	2.336	2.358	1.391	-1.611	-0.053
	$\pm$1.032	$\pm$0.178	$\pm$0.123	$\pm$0.175	$\pm$0.061
White Gaussian noise (SNR=10)					
$S.V._{\boldsymbol{c_e}}$	1.000	0.436	0.154	0.030	0.008
	$\pm$0.000	$\pm$0.014	$\pm$0.002	$\pm$0.000	$\pm$0.000
$S.V._{pcs}'$	1.000	0.465	0.264	0.103	0.078
	$\pm$0.000	$\pm$0.012	$\pm$0.004	$\pm$0.002	$\pm$0.001
c_{3y}	2.584	2.315	1.378	-1.561	-0.036
	$\pm$0.249	$\pm$0.114	$\pm$0.038	$\pm$0.043	$\pm$0.018
c_{3y}'	2.261	2.350	1.413	-1.628	0.042
	$\pm$1.126	$\pm$0.185	$\pm$0.154	$\pm$0.170	$\pm$0.076
Color Gaussian noise (SNR=10)					
$S.V._{\boldsymbol{c_e}}$	1.000	0.436	0.155	0.035	0.011
	$\pm$0.000	$\pm$0.015	$\pm$0.002	$\pm$0.000	$\pm$0.000
$S.V._{pcs}'$	1.000	0.467	0.270	0.103	0.081
	$\pm$0.000	$\pm$0.012	$\pm$0.004	$\pm$0.002	$\pm$0.001
c_{3y}	2.575	2.339	1.397	-1.519	-0.033
	$\pm$0.326	$\pm$0.129	$\pm$0.062	$\pm$0.070	$\pm$0.039
c_{3y}'	2.122	2.324	1.347	-1.584	-0.076
	$\pm$1.114	$\pm$0.253	$\pm$0.288	$\pm$0.248	$\pm$0.210

3.6 Conclusion

In this Chapter, we extend the co-prime theorem to PCS, and prove theorems about existence and seamless coverage of PCS for any index in HOS. We also demonstrate the unique existence of mapping scheme for deriving HOS from PCS. Based on these theorems, we propose the PCS-based HOS algorithm. One of its attractive advantages is the ability of significantly decreasing computational complexity while still maintaining the statistical properties.

We apply the PCS-based HOS algorithm in ARMA processes. The simulations indicate that this algorithm makes the computation less expensive, retains all properties of using HOS for order determination, and introduces new feature to double check the correctness of order selection. For the MA order selection, in high SNR environment ($SNR \geq 10$), it produces more variance in the zeroth lag, which does not affect the order decision, and in other places the variances remain within 10% of the amplitude. In low SNR cases ($SNR \approx 0$), it retains the amplitude estimations and introduces significant variance for order mismatch, which makes it better for the purpose of order selection. For the AR order determination, PCS makes minor difference after SVD approach.

CHAPTER 4

Co-prime sampling of HOS for channel estimation

4.1 Introduction

There are many communication problems that would benefit from using higher-order statistics (HOS). Because autocorrelation and power spectrum are phase blind, only Gaussian signals can be completely characterized by their first- and second-order statistics. It is of interest to utilize HOS in problems including 1) digital communication, such as matched filter for signal detection [17] and phase reconstruction [18], 2) identification of nonlinear systems [19], and nonminimum phase time-invariant channel models [20], and 3) non-Gaussian signal processing [21] and pattern recognition [22].

Broadly speaking, there are mainly two types of methods estimating HOS [21]: conventional nonparametric method based on Fourier transform, and parametric methods based on moving average (MA), autoregressive (AR), or autoregressive moving average (ARMA) models. The nonparametric approaches have better performance to quantify phase coupling at harmonically related frequency pairs, while the parametric methods are widely used owing to their computational efficiency and higher frequency resolution. The papers [21] and [23] provide further details about their properties. It is worth noting that both kinds of methods have high variance and require a large number of records to obtain smooth estimates, but increasing the number of segments is demanding on computation, and may increase bias and introduce nonstationarities [24].

The dilemma above was our motivation to design an algorithm estimating HOS via co-prime sampling, which had been proved useful in decreasing computational complexity while still maintaining statistical features of the original signal [25]. The basic idea of co-prime processing was to use Chinese reminder theorem with Bezout's identity to identify multiple frequencies from under-sampled sequences [26]. Vaidyanathan [27] further confined the rates of downsampling to be co-prime, and provided concrete demonstrations for sampling region so that co-prime sampled points were able to calculate all of the second order derivatives. Besides, the concept of coprimality was also used in system identification [28], filter banks design [27], and multidimensional and multirate signal processing [29]. However, co-prime signal processing had not been considered to derive HOS in previous research.

The main purpose of the Chapter **??** is to extend co-prime sampling calculating HOS and to develop a PCS-based HOS algorithm processing non-Gaussian signal or using in the nonminimum phase linear system. One of the advantages of the proposed algorithm is to introduce new trade-off for the implementation of HOS–the algorithm is able to maintain the same degree of complexity but to achieve better HOS estimates; or it reduces the complexity by several orders with a mild performance loss.

In Section 4.2 we describe the PCS-based HOS algorithm, analyze its complexity, and compare it with existing methods. The algorithm of channel parameter estimation is developed in Section 4.3, and its performance is evaluated in Section 4.4. Finally, we draw the conclusions in Section 4.5.

4.2 Estimate of cumulants in PCS and the complexity analysis

In this Section, we propose the PCS-based HOS algorithm and demonstrate how this algorithm reduces the computational complexity by several orders. In the following content, PCS $x_1[i], x_2[i], \ldots, x_r[i]$ denote the sampled sequences from in-

put signal related with pairwise relatively prime integers $n_1, n_2, \ldots, n_r$, respectively. Besides, two more conditions are assumed to simplify the derivation:

S.1 The input signals $\{x[n]\}$ are real, discrete, zero-mean, and wide-sense stationary to at least fourth order.

S.2 The segment length of the signal is equal to or longer than $2N$, where $N = \prod_{i=1}^{r} n_i$. This makes the implementation satisfying the Theorem 2 so that all indexes in HOS could be derived from the PCS.

Based on Theorems 1 and 2, given PCS $x[i]$, we are able to find any $\tau \in [0, N-1]$, such that

$$\tau = \sum_{i=1}^{r} a_i n_i, \quad a_i \in \left(-\frac{N}{n_i}, \frac{N}{n_i} \right). \tag{4.1}$$

The second-order autocorrelation ($r = 2$) in (3.6) is rewritten as

$$\hat{c}_{2x}(\tau) = \frac{1}{L} \sum_{l=0}^{L} x_1(a_1 + \frac{N}{n_1} l) x_2(a_2 + \frac{N}{n_2} l) \tag{4.2}$$

where L denotes the overall number of segmentations used for averaging. Because the resulting indexes in PCS are calculated via combination of different co-prime pairs rather than permutation of composite components, the estimates for theoretically symmetric points are not equal. Hence, they are further averaged within the segment

$$c_{2x}(\tau) = \frac{1}{2} \left(\hat{c}_{2x}(\tau) + \hat{c}_{2x}(-\tau) \right). \tag{4.3}$$

Similarly, the third- and four-order cumulants ($r = 3$ and $r = 4$) are estimated as

$$\hat{c}_{3x}(\tau_1, \tau_2) = \frac{1}{L} \sum_{l=0}^{L} \prod_{i=1}^{3} x_i(a_i + \frac{N}{n_i} l), \tag{4.4}$$

$$\begin{aligned} \hat{c}_{4x}(\tau_1, \tau_2, \tau_3) = & \frac{1}{L} \sum_{l=0}^{L} \prod_{i=1}^{4} x_i(a_i + \frac{N}{n_i} l) - \hat{c}_{2x}(\tau_1) \hat{c}_{2x}(\tau_2 - \tau_3) \\ & - \hat{c}_{2x}(\tau_2) \hat{c}_{2x}(\tau_3 - \tau_1) - \hat{c}_{2x}(\tau_3) \hat{c}_{2x}(\tau_1 - \tau_2), \end{aligned} \tag{4.5}$$

and further averaged via symmetry property [23]

$$\begin{aligned}\hat{c}_{3x}(\tau_1,\tau_2) =& \frac{1}{6}\left[\hat{c}_{3x}(\tau_1,\tau_2)+\hat{c}_{3x}(\tau_2,\tau_1)\right.\\ &+\hat{c}_{3x}(-\tau_2,\tau_1-\tau_2)+\hat{c}_{3x}(-\tau_1,\tau_2-\tau_1)\\ &\left.+\hat{c}_{3x}(\tau_2-\tau_1,-\tau_1)+\hat{c}_{3x}(\tau_1-\tau_2,-\tau_2)\right]. \end{aligned} \tag{4.6}$$

The fourth-order cumulant has twenty-four symmetric parts which are also worth exploring in the computation process.

The following content shows that the complexity improvement is nontrivial via estimating the HOS based on PCS. To quantify the complexity, we use the "Big O" notation to denote the asymptotic limitation of a function when the argument increases towards infinity. Both the parametric and nonparametric method have to estimate cumulant based on expectation, which is the first part of the complexity. It mainly comes from multiplication traversal in (3.1), (3.2), and (3.3). The operations must be performed record by record which in turn causes the complexity to increase exponentially. To estimate kth-order cumulant, the computation complexity is $O(N^k)$ where N is the length of segment.

$$\begin{aligned}S_{2,\tau_1,(\tau_2-\tau_3)}(\omega_1,&\omega_2,\omega_3) \triangleq \mathcal{F}[c_{2x}(\tau_1)c_{2x}(\tau_2-\tau_3)]\\ &=c_{2x}(\tau_1)c_{2x}(\tau_2-\tau_3)exp\left(-j2\pi\sum_{i=1}^{3}\tau_i\omega_i\right)\\ &=\left[\sum_{(a_1-a_2)=0}^{N}x_1(a_1)x_2(a_2)\sum_{(a_3-a_4)=0}^{N}x_3(a_3)x_4(a_4)\right]exp\left(-j2\pi\sum_{i=1}^{3}\tau_i\omega_i\right)\\ &=\sum_{(a_1-a_2)=0}^{N}x_1(a_1)exp\left(\frac{-j2\pi a_1(\omega_1-\omega_2-\omega_3)}{N}\right)x_2(a_2)exp\left(\frac{-j2\pi a_1\omega_1}{N}\right)\\ &\quad\sum_{(a_3-a_4)=0}^{N}x_3(a_3)exp\left(\frac{-j2\pi a_3\omega_2}{N}\right)x_4(a_4)exp\left(\frac{-j2\pi a_4\omega_3}{N}\right)\end{aligned} \tag{4.7}$$

The nonparametric method further requires frequency averaging which makes it more impractical for longer sequences. The classical FFT algorithm achieves computational complexity $O(NlogN)$ for second-order spectrum and $O(N^{r-1}logN)$ for the rth-order Fourier transform [31]. Including the calculation of cumulants, the overall complexity achieves as high as $O(N^{2k-1}logN)$.

On the other hand, PCS is able to decrease the calculation of cumulant by at least one order. Suppose the length of segmentation is $2N$. The complexity for third-order PCS cumulant is

$$\frac{2N}{n_1} \times \frac{2N}{n_2} \times \frac{2N}{n_3} \sim O(N^2) \tag{4.8}$$

where $N = \prod_{i=1}^{3} n_i$.

For nonparametric approach, the optimization is more significant by making the Fourier transform in-place without complicated index mapping. The following analysis is inspired by the structure of prime-factor FFT in [33]. For the kth-order spectrum, we consider (3.4) to be a series of two-dimensional Fourier transform. Then, the formula is rewritten as

$$\begin{aligned} S_{kx}(\omega_1, \omega_2, \ldots, \omega_{k-1}) = \frac{1}{L} \sum_{l=1}^{L} \sum_{\tau_{k-1}=-N}^{N} \Bigg[\ldots \Bigg[\sum_{\tau_1=-N}^{N} c_{kx}(\tau_1, \tau_2, \ldots, \tau_{k-1}) \\ exp\left(\frac{-j2\pi\omega_1\tau_1}{N}\right)\Bigg] \ldots \Bigg] exp\left(\frac{-j2\pi\omega_{k-1}\tau_{k-1}}{N}\right). \end{aligned} \tag{4.9}$$

Specifically, we can simplify the bispectrum as

$$\begin{aligned} S_{3x}(\omega_1, \omega_2) = \sum_{(a_1-a_3)=-N}^{N} \Bigg[\sum_{(a_1-a_2)=-N}^{N} x_1(a_1) exp\left(\frac{-j2\pi a_1(\omega_1+\omega_2)}{N}\right) \\ x_2(a_2) exp\left(\frac{-j2\pi a_2\omega_1}{N}\right)\Bigg] x_3(a_3) exp\left(\frac{-j2\pi a_3\omega_2}{N}\right). \end{aligned} \tag{4.10}$$

The averaging among L segments is deliberately omitted for simpler formula abstraction. It can be easily restored in the implementation. Because L increases

linearly along with the length of signal, it is negligible compared with the exponentially increased asymptotic property of N. In (5.11), complex multiplications are in-place calculated, which only change a constant factor of the asymptotic limitation. Hence, the calculation procedure decreased from $O(N^5 logN)$ of the existing HOS to quadratic form $O(N^2)$ for PCS-based HOS.

The fundamental difference between (5.12) and (3.4) is that without PCS, c_{kx} is calculated by the exhausted multiplying among N records for k times, which cannot be decomposed. Furthermore, according to Corollary 2, there is one unique mapping for a given set of PCS. So, this process can be deconvoluted for in-place Fourier transform.

The analysis for trispectrum is more complex because of cross terms of second-order cumulants in (4.5). We split its Fourier transform into four parts as

$$S_{4x}(\omega_1, \omega_2, \omega_3) = S_{4,\tau_1,\tau_2,\tau_3} - S_{2,\tau_1,(\tau_2-\tau_3)} - S_{2,\tau_2,(\tau_3-\tau_1)} - S_{2,\tau_3,(\tau_1-\tau_2)}. \tag{4.11}$$

The first part is the fourth-order moment. Similar to the third-order cumulant, its complexity is $O(N^3)$. The other three cross terms are equivalent in complexity because they are permutations of PCS $x_1(a_1)$, $x_2(a_2)$, $x_3(a_3)$ and $x_4(a_4)$. Take the second term as example in (4.7). It can be regarded as products of in-place second-order Fourier transform with complexity $O(N)$. The overall complexity of $S_{2,\tau_1,(\tau_2-\tau_3)}$ is $O(N^2)$. Hence, the trispectrum is added by one multiplicative term with complexity $O(N^3)$ and three cross terms of second-order spectrum with complexity $O(N^2)$. The resulting overall complexity is decreased from $O(N^7 logN)$ to $O(N^3)$ for PCS.

4.3 MA channel estimation algorithm using HOS

As an application, we implement PCS-HOS to estimate impulse responses of MA system. In [20], it demonstrated that MA coefficients could be computed recur-

sively in closed form using both correlation and third-order diagonally sliced cumulant $c_{3,y}(\tau, \tau)$. The paper also justified the uniqueness of the solution in a least-square form. In [35], the algorithm was extended to the fourth order, and further improved using more slices of time increments. However, as analyzed before, the computation increases cubically for third-order cumulant along with the length of segment. It might render the algorithm impractical before the estimate variation converges to an optimal state. In the following content, we propose a modification of existing least-square approach and use overdetermined formulas to suppress the additional estimate variance caused by PCS.

Consider the finite impulse response (FIR) signal model:

$$x(t) = \sum_{i=0}^{q} h(i)w(t-i) \tag{4.12}$$

where the observations consist both the signal $x(t)$ and zero-mean noise $v(t)$:

$$y(t) = x(t) + v(t). \tag{4.13}$$

We have three assumptions for the following derivation:

S.1 The driving noise sequence $w(t)$ is sufficiently long, non-Gaussian, zero mean and independent and identical distributed (i.i.d.). The statistical characteristics includes $E[w^2(t)] = \sigma_w^2$, $E[w^k(t)] = \gamma_{kw}$, and $E[w^{2k}(t)] < +\infty$, for $k = 3$ or 4.

S.2 The measurement noise sequence $v(t)$ is assumed to be zero mean, but does not necessarily to be Gaussian. It is also i.i.d. and independent of $w(t)$ with $E[v^2(t)] = \sigma_v^2$, $E[v^k(t)] = \gamma_{kv}$, and $E[v^{2k}(t)] < +\infty$, for $k = 3$ or 4.

S.3 The MA order q is known. For $|i| > q$, $h(i) = 0$. Hence, $x(i) = 0$ and $y(i) = 0$.

4.3.1 MA system identification using 3rd-order cumulant

We extend the definition (3.2) as follow:

$$c_{3x}(\tau_1, \tau_2) = \gamma_{3w} \sum_{k=0}^{q} h(k)h(k+\tau_1)h(k+\tau_2) = \gamma_{3w} \sum_{k=0}^{q} h(k)g(k; \tau_1, \tau_2) \tag{4.14}$$

where

$$g(k; \tau_1, \tau_2) = h(k+\tau_1)h(k+\tau_2). \tag{4.15}$$

Let $G(z; \tau_1, \tau_2)$ denote the z-transform of the sequence $g(k; \tau_1, \tau_2)$, then we can have

$$C_{3x}(z; \tau_1, \tau_2) = \gamma_{3w} H(z^{-1}) G(z; \tau_1, \tau_2) = \gamma_{3w} z^{\tau_2 - \tau_1} H(z^{-1})[H(z) * H(z)]. \tag{4.16}$$

If we fix τ_1 and replace τ_2 with a different time increments τ_3, we could have another 3rd-order cumulant containing $\gamma_{3w} H(z^{-1})$.

$$\gamma_{3w} H(z^{-1}) = C_{3x}(z; \tau_1, \tau_2)[G(z; \tau_1, \tau_2)]^{-1} = C_{3x}(z; \tau_1, \tau_3)[G(z; \tau_1, \tau_3)]^{-1}. \tag{4.17}$$

The two terms in the right hand side could form the equation

$$\begin{aligned} &\sum_{k=0}^{q} h(k+\tau_1)h(k+\tau_2)c_{3x}(k-\tau_1, \tau_3) \\ &\qquad = \sum_{k=0}^{q} h(k+\tau_1)h(k+\tau_3)c_{3x}(k-\tau_1, \tau_2). \end{aligned} \tag{4.18}$$

Let $\tau_2 - \tau_1 = 0$, $\tau_3 - \tau_1 = q$ and $h(0) = 1$, and denote τ_1 as τ. Based on the assumption (S.2) and (S.3), we can have the formula relating impulse response coefficients $h(t)$ with output $y(t)$ and measurement noise $v(t)$

$$h(q)[c_{3y}(-\tau, 0) - \gamma_{3v}\delta(\tau)] = c_{3y}(-\tau, q) + \sum_{i=1}^{q} h^2(i)c_{3y}(i-\tau, q). \tag{4.19}$$

Another connection between $h(t)$ and $y(t)$ comes from both second-order and third-order. The autocorrelation coefficients of output signal satisfy

$$c_{2x}(\tau) = E\left[x(k)x(k+\tau)\right] = \sigma^2 \sum_{k=max(0,-m)}^{min(q,q-k)} h(k)h(k+\tau) \tag{4.20}$$

where $k \in [-q, q]$ for all k. Its z transform of is

$$C_{2x}(z;\tau) = \sigma_w^2 H(z^{-1})H(z). \tag{4.21}$$

Combining (4.21) with (4.14) and letting τ in (4.21) equal to τ_1 in (4.14), we could have

$$G(z;\tau_1,\tau_2)C_{2x}(z;\tau_1) = \frac{\sigma_w^2}{\gamma_{3w}} H(z)C_{3x}(z;\tau_1,\tau_2) \tag{4.22}$$

$$\sum_{k=0}^{q} h(k)h(k+\tau_2-\tau_1)r_x(\tau_1-k) = \frac{\sigma_w^2}{\gamma_{3w}} \sum_{k=0}^{q} h(k)c_{3x}(\tau_1-k,\tau_2-k). \tag{4.23}$$

Then, if we let $\tau_1 = \tau_2$, replace both symbols as τ, and let $\varepsilon = \sigma_w^2/\gamma_{3w}$, we have

$$\begin{aligned}
&[r_y(\tau) - \sigma_v^2\delta(\tau)] + \sum_{k=1}^{q} h^2(k)[r_y(\tau-k) - \sigma_v^2\delta(\tau-k)] \\
&\qquad = \varepsilon[c_{3y}(\tau,\tau) - \gamma_{3v}\delta(\tau)] + \sum_{k=1}^{q}[\varepsilon h(k)][c_{3y}(\tau-k,\tau-k) - \gamma_{3v}\delta(\tau-k)] \\
&\qquad = \varepsilon[c_{3y}(-\tau,0) - \gamma_{3v}\delta(\tau)] + \sum_{k=1}^{q}[\varepsilon h(k)][c_{3y}(k-\tau,0) - \gamma_{3v}\delta(k-\tau)].
\end{aligned} \tag{4.24}$$

Let $1 \le |\tau| \le q$ in (4.18), and $-q \le \tau \le -1$ and $q+1 \le \tau \le 2q$ in (4.24). We have $4q$ equations for $2q+2$ unknown variables, which could be solved to get least-square solution of the impulse response coefficients. The compact form is

$$\boldsymbol{Mh} = \boldsymbol{c} \tag{4.25}$$

where

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M_1} \\ \boldsymbol{M_2} \end{bmatrix}$$

$$\boldsymbol{h} = [\varepsilon\ \varepsilon h(1)\ \varepsilon h(2) \dots \varepsilon h(q)\ h^2(1) \dots h^2(q)\ h(q)]^T$$

$$\boldsymbol{c} = [\boldsymbol{c_1}\ \boldsymbol{c_2}]^T$$

$$\boldsymbol{c_1} = [r_y(-q)\ r_y(-q+1) \dots r_y(-1)\ 0\ 0 \dots 0]$$

$$\boldsymbol{c_2} = [c_{3y}(q,q)\ c_{3y}(q-1,q) \dots c_{3y}(1,q)\ 0\ 0 \dots 0].$$

The matrix representations of $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ are described in (4.26) and (4.27).

$$\boldsymbol{M_1} = \begin{bmatrix} c_{3y}(q,0) & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ c_{3y}(q-1,0) & c_{3y}(q,0) & \dots & 0 & 0 & -r_y(-q) & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{3y}(1,0) & c_{3y}(2,0) & \dots & c_{3y}(q,0) & 0 & -r_y(-2) & -r_y(-3) & \dots & 0 & 0 \\ 0 & c_{3y}(-q,0) & \dots & c_{3y}(-2,0) & c_{3y}(-1,0) & -r_y(q) & -r_y(q-1) & \dots & -r_y(-1) & 0 \\ 0 & 0 & \dots & c_{3y}(-3,0) & c_{3y}(-2,0) & 0 & -r_y(q) & \dots & -r_y(-2) & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & c_{3y}(-q,0) & 0 & 0 & \dots & -r_y(q) & 0 \end{bmatrix} \tag{4.26}$$

$$\boldsymbol{M_2} = \begin{bmatrix} 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & c_{3y}(q,0) \\ 0 & \dots & 0 & -c_{3y}(q,q) & 0 & \dots & 0 & 0 & c_{3y}(q-1,0) \\ 0 & \dots & 0 & -c_{3y}(q-1,q) & -c_{3y}(q,q) & \dots & 0 & 0 & c_{3y}(q-2,0) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -c_{3y}(2,q) & -c_{3y}(3,q) & \dots & -c_{3y}(q,q) & 0 & c_{3y}(1,0) \\ 0 & \dots & 0 & -c_{3y}(0,q) & -c_{3y}(1,q) & \dots & -c_{3y}(q-2,q) & -c_{3y}(q-1,q) & c_{3y}(-1,0) \\ 0 & \dots & 0 & 0 & -c_{3y}(0,q) & \dots & -c_{3y}(q-3,q) & -c_{3y}(q-2,q) & c_{3y}(-2,0) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & -c_{3y}(0,q) & c_{3y}(-q,0) \end{bmatrix} \tag{4.27}$$

Finally, the least square solution in compact form is

$$\hat{\boldsymbol{h}}_{\boldsymbol{LS}} = (\boldsymbol{M}^{\boldsymbol{T}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\boldsymbol{T}}\boldsymbol{c}. \tag{4.28}$$

There are several aspects worth mentioning in both deriving (4.18) and (4.24) and transforming them into matrices representation.

Remark 1. *The impulse response $\delta(\tau)$ in (4.18) and (4.24) is based on the assumption S.2, which makes the cross-correlation $E[x(t)v(t)] = 0$ for all t, and $E[v(t_1)v(t_2)] = 0$ for $t_1 \neq t_2$.*

Remark 2. *In deriving the formula (4.24), we use the symmetric property of 3rd-order cumulants. Specifically, $c_{3y}(m, m) = c_{3y}(-m, 0)$ simplify the slicing for 3rd-order cumulant only taking place for $c_{3y}(i, 0)$ and $c_{3y}(i, q)$ for $1 \leq |i| \leq q$.*

Remark 3. $\boldsymbol{M}$ *is a $4q \times (2q + 2)$ matrix, which can be further divided into two $2q \times (2q + 2)$ matrices $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$. They are derived from (4.18) and (4.24) respectively.* $\boldsymbol{h}$ *is a $2q + 2$ column vector.* $\boldsymbol{c}$ *is a $4q$ column vector, which is evenly divided into two sub-column vectors, $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$, which also come from (4.18) and (4.24).*

Remark 4. *In order to eliminate the effect of measurement noise, we exclude the rows coming from $\tau = 0$ or $k = \tau$ when building matrices.*

Remark 5. *There are many zeros elements in $\boldsymbol{M}$ and $\boldsymbol{c}$, because for the FIR system we assume $b(i) \neq 0$ only for $-q \leq i \leq q$. That is, $c_{3y}(i, j) \neq 0$ and $r_y(i) \neq 0$ for $m, n \in [-q, q]$.*

4.3.2 MA system identification using 4th-order cumulant

When the input signal $w(t)$ has symmetric probability density function, the third-order cumulants are identically zero. Examples of such situation include Bernoulli-Gaussian random variables considered in [32], and signal with Laplace distribution in [20]. In this circumstance, it is necessary to exploit the fourth-order cumulant to fully utilize its non-Gaussian properties.

The general form of z-transform of the kth-order cumulant is

$$C_{kx}(z) = \gamma_{kw} H(z^{-1}) G(z; \tau_1, \ldots, \tau_{k-1}) \tag{4.29}$$

where

$$G(z;\tau_1,\ldots,\tau_{k-1}) = z^{\tau_{k-1}-\tau_1}\underbrace{H(z)*H(z)\cdots*H(z)}_{(k-1)\text{ times.}} \tag{4.30}$$

Similarly to the third-order case, we use the same assumptions stated at the beginning of this section to derive the equations using autocorrelations and 4th-order cumulants. The expectation of input cumulant is derived from the definition (3.3) and (4.12)

$$c_{4x}(\tau_1,\tau_2,\tau_3) = \gamma_{4w}\sum_{k=0}^{q} h(k)h(k+\tau_1)h(k+\tau_2)h(k+\tau_3). \tag{4.31}$$

After Z-transform, let $\tau_1 = \tau$, $\tau_2 - \tau = m$, $\tau_3 - \tau = n$, we have the following relation:

$$\begin{aligned} C_{4,x}(z;\tau,\tau+m,\tau+n) &= \gamma_{4w}H(z^{-1})G(z;\tau,\tau+m,\tau+n) \\ &= \gamma_{4w}z^m z^n H(z^{-1})[H(z)*H(z)*H(z)]. \end{aligned} \tag{4.32}$$

The counterparts of (4.18) and (4.19) are derived for the same first-step time increment τ with two different groups of 2nd- and 3rd-step increments m_0, n_0 and m_1, n_1

$$\begin{aligned} &\sum_{k=0}^{q} h(k)h(k+m_1)h(k+n_1)c_{4x}(k-\tau,m_0,n_0) \\ &\qquad = \sum_{k=0}^{q} h(k)h(k+m_0)h(k+n_0)c_{4x}(k-\tau,m_1,n_1). \end{aligned} \tag{4.33}$$

Let $m_0 = n_0 = 0$ and $m_1 = n_1 = q$, we obtain the first formula constructing the matrix for least-square solution:

$$h^2(q)[c_{4y}(-\tau,0,0) - \gamma_{4v}\delta(\tau)] = c_{4y}(-\tau,q,q) + \sum_{k=1}^{q} h^3(k)c_{4y}(k-\tau,q,q). \tag{4.34}$$

With consistent notion for letting $\tau_2 - \tau = m$, $\tau_3 - \tau = n$, the relation between autocorrelation and 4th-order cumulant is

$$\sum_{k=0}^{q} h(k)h(k+m)h(k+n)r_x(\tau-k)$$
$$= \varepsilon \sum_{k=0}^{q} h(k)c_{4x}(\tau-k, \tau+m-k, \tau+n-k). \tag{4.35}$$

Let $m = n = 0$, we have the second formula for the matrix of least-square solution

$$[r_y(\tau) - \sigma_v^2\delta(\tau)] + \sum_{k=1}^{q} h^3(k)[r_y(\tau-k) - \sigma_v^2\delta(\tau-k)]$$
$$= \varepsilon[c_{4y}(-\tau, 0, 0) - \gamma_{4v}\delta(\tau)] + \sum_{k=0}^{q}[\varepsilon h(k)][c_{4y}(k-\tau, 0, 0) - \gamma_{4v}\delta(\tau-k)]. \tag{4.36}$$

Let $1 \leq |\tau| \leq q$ in (4.34), and $-q \leq \tau \leq -1$ and $q+1 \leq \tau \leq 2q$ in (4.36). The compact form is

$$\boldsymbol{Mh} = \boldsymbol{c} \tag{4.37}$$

where

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M_1} \\ \boldsymbol{M_2} \end{bmatrix}$$
$$\boldsymbol{h} = [\varepsilon\ \varepsilon h(1)\ \varepsilon h(2) \ldots \varepsilon h(q)\ h^3(1) \ldots h^3(q)\ h^2(q)]^T$$
$$\boldsymbol{c} = [\boldsymbol{c_1}\ \boldsymbol{c_2}]^T$$
$$\boldsymbol{c_1} = [r_y(-q)\ r_y(-q+1) \ldots r_y(-1)\ 0\ 0 \ldots 0]$$
$$\boldsymbol{c_2} = [c_{4y}(q,q,q)\ c_{4y}(q-1,q,q) \ldots c_{4y}(1,q,q)\ 0\ 0 \ldots 0].$$

The dimensionalities for $\boldsymbol{M}$, $\boldsymbol{h}$, and $\boldsymbol{c}$ are the same as using 3rd-order cumulants. The structure of matrices of $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ could also refer to (4.26, 4.27) and replace $c_{3y}(\tau, 0)$ and $c_{3y}(\tau, q)$ with $c_{4y}(\tau, 0, 0)$ and $c_{4y}(\tau, q, q)$.

4.4 Simulation results

4.4.1 Basic setting

In order to fully explore properties of PCS, we apply it to estimate the simplified LTE spacial channel model [34] where the channels are assumed to be MA models with six paths. Furthermore, we assume each path has only one sub-path, and the channel characteristics remain invariant over the processed data symbols, which means the channel is slow fading for the very high signaling rates. MA(2) and MA(5) are simulated with the inputs and outputs:

$$x_1(t) = w(t) - 2.333w(t-1) + 0.667w(t-2) \tag{4.38}$$

$$\begin{aligned} x_2(t) =& w(t) + 0,1w(t-1) - 1.87w(t-2) \\ &+ 3.02w(t-3) - 1.435w(t-4) \\ &+ 0.49w(t-5) \end{aligned} \tag{4.39}$$

$$y_1(t) = x_1(t) \tag{4.40}$$

$$y_2(t) = x_2(t). \tag{4.41}$$

Since the estimation is more susceptible for abnormal samples due to the sparse sampling, it is of interest to see how PCS performs in computing higher-order cumulants. In both model, we assume $SNR = +\infty$. For channel estimation, we generate signal sequence with 700000 sample points for a single run and average results from 50 times of Monte Carlo simulations for every scenario. Besides, we do not use any filter to preprocess $w(t)$, and set the overlap percentage equal to zero. We also conservatively select the PCS by setting $n_1 = 1$, $n_2 = 2$, and $n_3 = 3$ for PCS-HOS(3), and $n_1 = 1$, $n_2 = 1$, $n_3 = 2$, and $n_4 = 3$ for PCS-HOS(4). Because the purpose of this paper is to indicate the validity of PCS-HOS algorithm, choosing smaller PCS

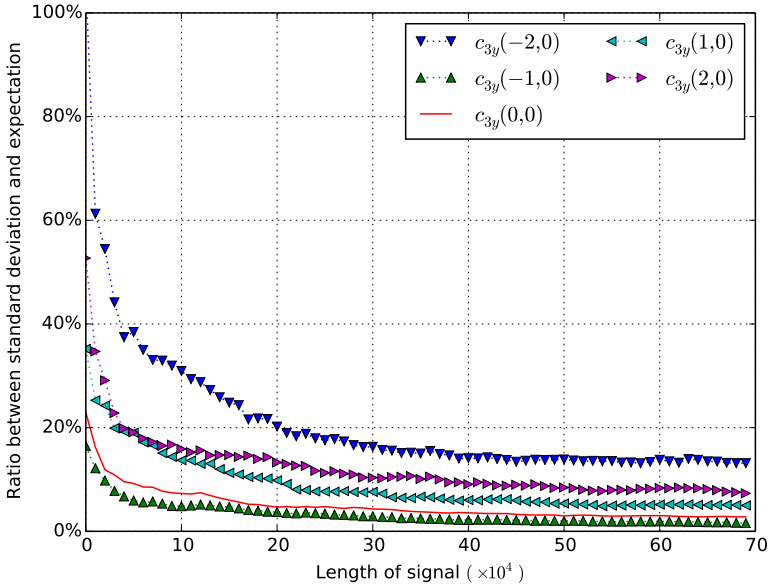

Figure 4.1: $y_1[n]$ cumulant variance using PCS-HOS(3)

with less variance is preferable. Concerning how to accelerate convergence is beyond the scope of this paper.

4.4.2 Convergence of cummulant calculated by PCS-based HOS

Because the absolute value of MA coefficients vary from 0.1 to 3.02, the variance itself cannot fully reveal how well the estimation is. We use the ratio between standard deviation and its expectation as vertical axis to indicate the convergence.

Figures 4.1 and 4.2 show the convergence of third-order PCS for estimating cumulants with 2 taps or 5 taps. The estimate variances fall below 40% of the expected value after $200K$ samples for MA(2), and 250K samples for MA(5). The only exception is $c_{3y}(-5,0)$. This is because there are less PCS points to average for larger difference. The potential improvements are filter the data before calculating cumulants, or use overlap windows to increase the candidates for averaging.

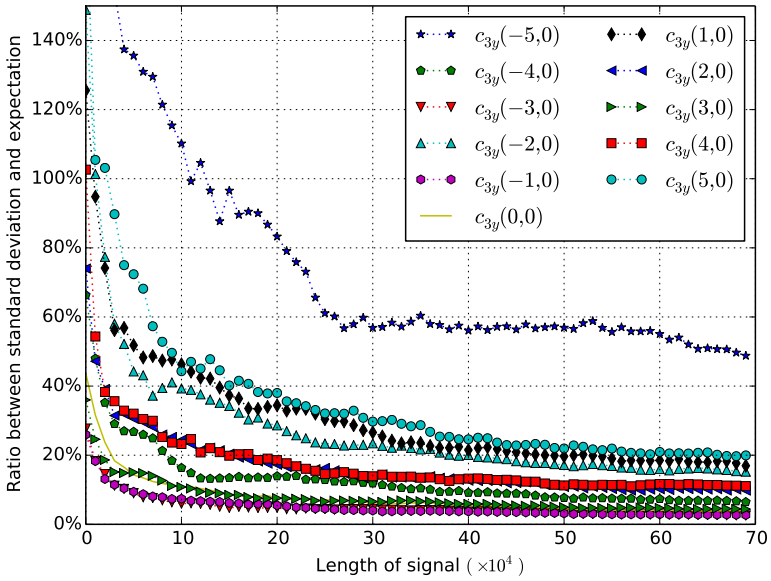

Figure 4.2: $y_2[n]$ cumulant variance using PCS-HOS(3)

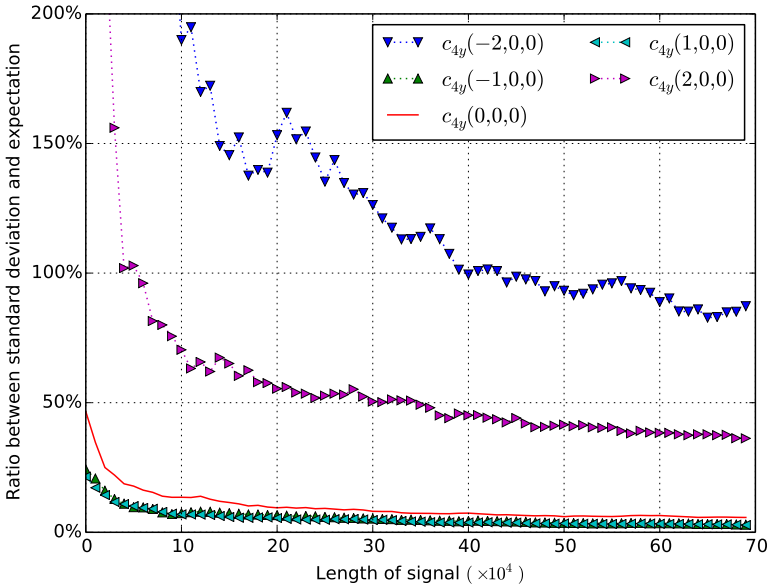

Figure 4.3: $y_1[n]$ cumulant variance using PCS-HOS(4)

Figures 4.3 and 4.4 indicate the convergence of PCS-HOS(4). Compared with Figures 4.1 and 4.2, they dramatically increase the estimation variance and decrease

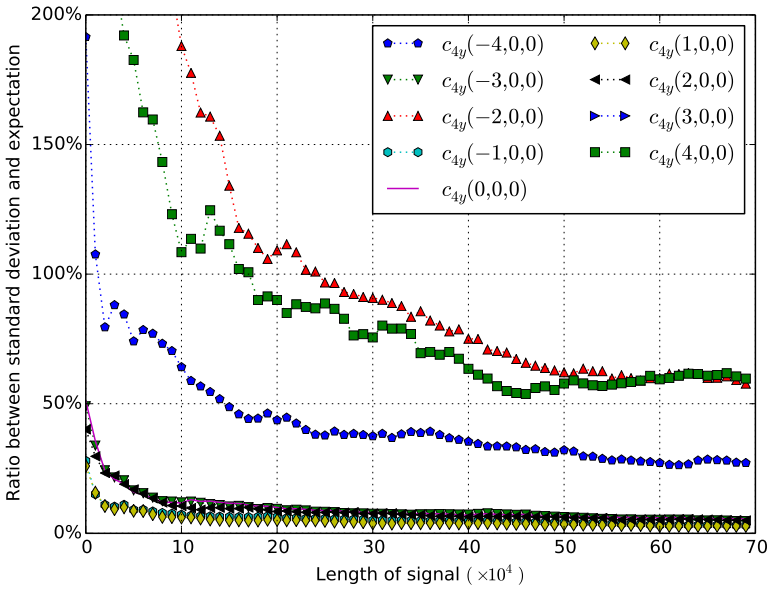

Figure 4.4: $y_2[n]$ cumulant variance using PCS-HOS(4)

the speed of convergence. Besides, Figure 4.4 is able to include only four out of five taps, because if we include $c_{4y}(-5,0,0)$ and $c_{4y}(5,0,0)$, their large values would render the other curves indistinguishable.

4.4.3 Convergence of MA parameter estimation

Figures 4.5 and 4.6 show the performance of estimating coefficients for MA(2) and MA(5) using PCS-HOS(3). Comparing to Figures 4.1 and 4.2, the performance of MA coefficients estimation correlates with the convergence of cumulant estimation. After 200k sample points, output results of MA(2) are fairly stable as the variance of cumulant estimations falls into 40% range of the expected values. Although the 1st tap of the MA(5) system seems abnormally unstable in Figure 4.6, it is mainly because the real value of that coefficient is very small (0.100). The general estimates are also stable after 200k samples.

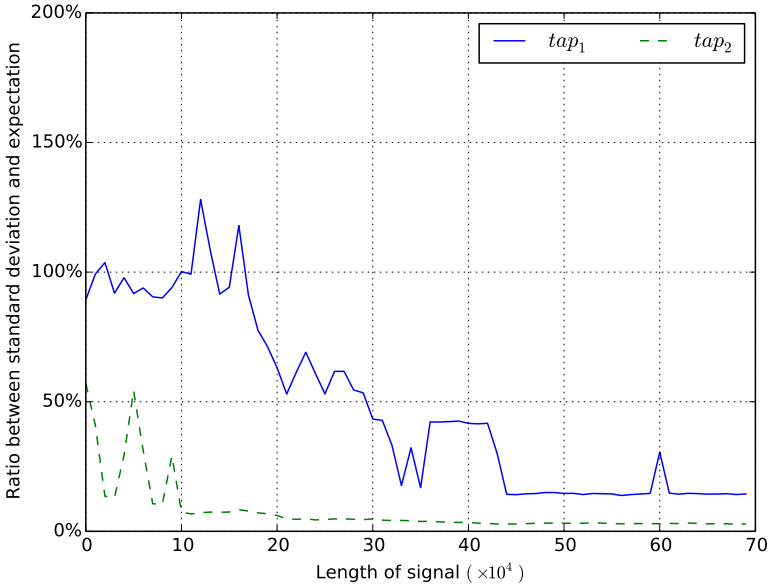

Figure 4.5: Variance of MA(2) using PCS-HOS(3)

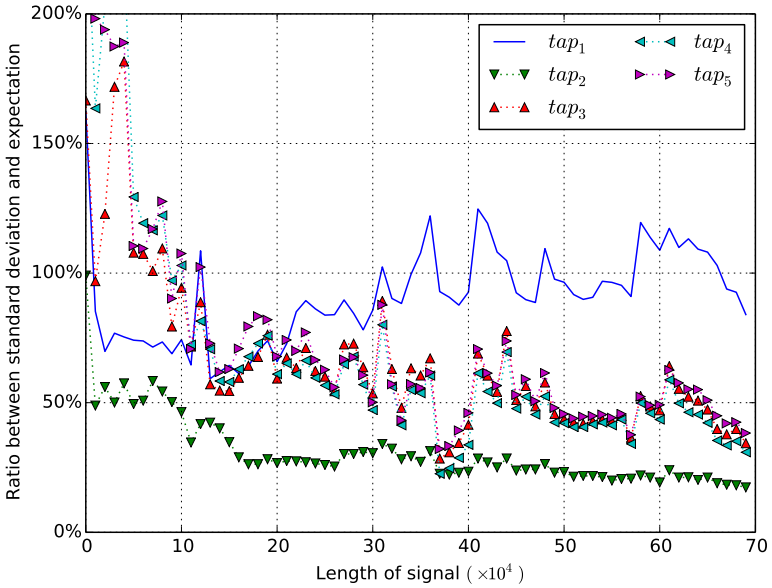

Figure 4.6: Variance of MA(5) using PCS-HOS(3)

The performances of PCS-HOS(4) in Figures 4.7 and 4.8 indicate that even for the most conservative choice of PCS, it is still impractical to use it combining with

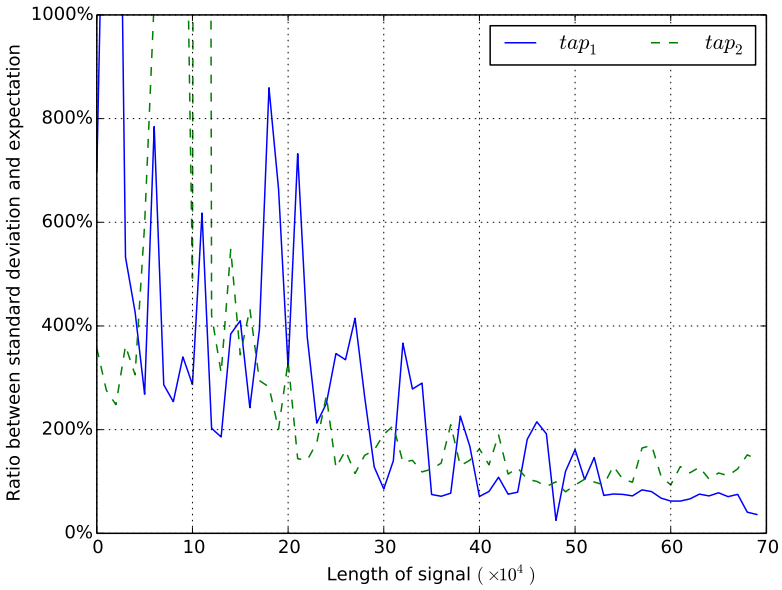

Figure 4.7: Variance of MA(2) using PCS-HOS(4)

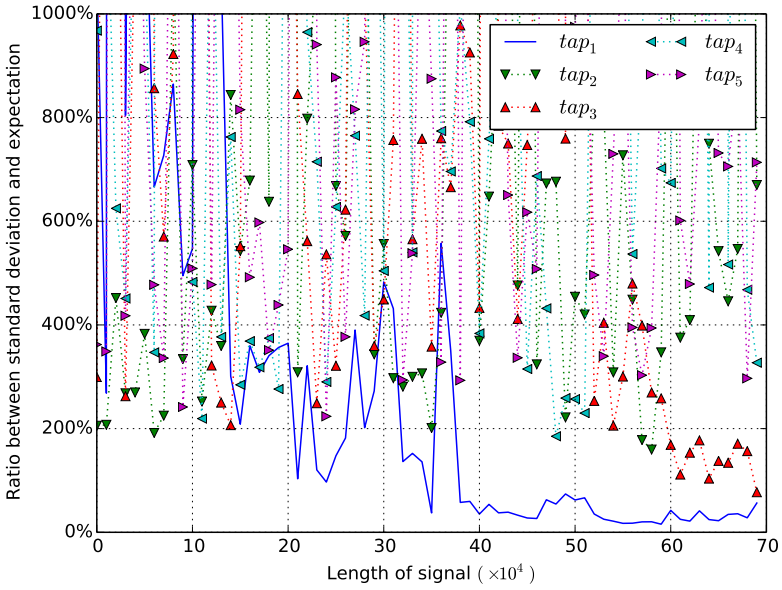

Figure 4.8: Variance of MA(5) using PCS-HOS(4)

HOS for system identification. Although PCS-HOS(4) performs relatively good for MA(2) system comparing to its performance in MA(5), the ratio between standard

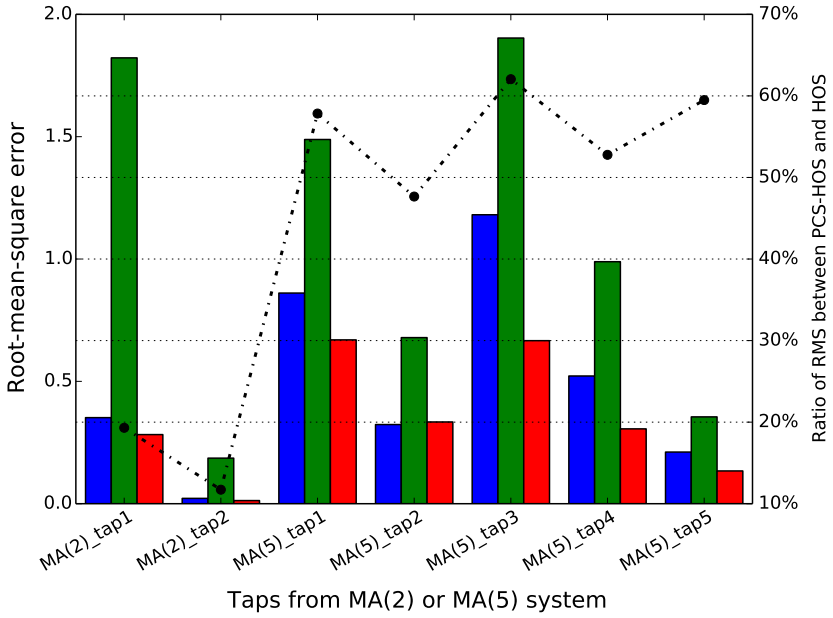

Figure 4.9: Performance comparison between PCS-HOS(3) and original HOS

deviation and expected value has amplified four times compared with the corresponding curves of using PCS-HOS(3) in Figure 4.5. In Figure 4.8, expect the first tap, all other taps remain unstable after 700k sample points. Referring to matrix representation of least-square solution (4.26, 4.27), it needs cumulants with all delays to estimate every tap in MA system. Because PCS-HOS(4) cannot confine the cumulants $c_{4y}(-5,0,0)$, $c_{4y}(-5,-5,-5)$, $c_{4y}(5,0,0)$, and $c_{4y}(5,5,5)$, MA(5) cannot get a stable estimate. The numeric results in Table 4.1 further reveal that the unstable cumulant estimations result in the bias for the parameter estimates.

4.4.4 Performance comparison

In Figure 4.9, the performances are quantified via root-mean square error (RMS). Three candidates in the comparison are all third-order HOS, including PCS-HOS(3) using $700k$ sample points in the left side (blue), existing HOS using $110k$ samples in the middle (green), and existing HOS using $700k$ in the right side (red). The reason

for choosing the HOS with $110k$ samples it because it has the same computational complexity as the PCS-HOS(3). The left vertical axis indicates the values of RMS, and the right vertical axis is the ratio of performance gain between PCS-HOS and he existing HOS with the same complexity. We can see that in MA(2) system PCS-HOS(3) decreases RMS for more than 80%. For MA(5) system, the performance gain is less significant, but it is still able to have about 45% performance gain on average. Nevertheless, it is worth mentioning that the performance gain and complexity simplification are acquired at the expense of averaging longer sequence, which increases the delay of the system. For the same length of signal sequences, existing HOS has roughly six times more complexity than PCS-HOS(3), and outperforms 12% less RMS in MA(2) system and 26% less RMS in MA(5). There are obvious differences for the last three taps of the MA(5) system. It is because the larger the tap of MA is, the wider range of cumulants it needs to estimate the coefficient. Consequently, the larger variance for longer cumulants in Figure 4.2 deteriorate the performance of PCS-HOS.

In Table 4.1, the proposed algorithm are compared with existing benchmarks. Both the results of GM and Tug come from the paper [35]. Note that these results are summarized from twenty-five Monte Carlo simulations, and the length of sequence in the simulation is 5120. But they are still comparable, since all results are statistical summarization and our purpose is to prove the validity of PCS-HOS. The "PCS3" and "PCS4" indicate PCS-HOS(3) and PCS-HOS(4) algorithms. The statistics of each algorithm include arithmetic mean, standard deviation, and RMS. For the benchmark algorithms, the "Tug" is much better than the "GM". "PCS3" outperforms "Tug" in estimating all of the five taps. Although "PCS4" has smaller standard deviation and RMS than the "GM", but after the second tap, the estimates bias renders the results useless.

Table 4.1: Parameter Estimates

Algo.	Taps	h(1)	h(2)	h(3)	h(4)	h(5)
	Val.	0.100	-1.870	3.020	-1.435	0.490
GM	mean	18.654	-39.996	23.025	-4.804	1.056
	std.	81.604	177.375	112.094	31.080	9.414
	RMS	83.687	181.409	114.233	31.262	9.431
Tug	mean	0.366	-0.626	0.960	-0.765	0.311
	std.	1.503	1.087	1.125	0.921	0.365
	RMS	1.526	1.652	2.347	1.139	0.407
PCS3	mean	0.716	-1.763	2.080	-1.018	0.317
	std.	0.601	0.305	0.715	0.315	0.122
	RMS	0.860	0.323	1.180	0.521	0.211
PCS4	mean	1.043	-0.490	-0.472	0.100	0.045
	std.	0.589	3.280	0.364	0.326	0.326
	RMS	1.112	3.558	3.511	1.569	0.551

4.5 Conclusions

In order to use the co-prime signal processing HOS, we extend the co-prime theorem to PCS, and prove theorems about existence and seamless coverage of PCS for any index in HOS. We also demonstrate the unique existence of mapping scheme for deriving HOS from PCS. Based on these theorems, we propose the PCS-HOS algorithm. This algorithm significantly alleviates the computational demands for calculating HOS while still maintains the statistical properties. If the system has no constraint on memory or concurrency, the "in-place" characteristics of PCS-HOS can be further accelerated based on parallel computing.

We apply the PCS-HOS algorithm to channel estimation where MA models are assumed. In MA(2) channel, the PCS-HOS(3) is able to achieve 80% performance gain comparing to existing HOS algorithm with the same computational complexity, or 12% performance loss using only 15% complexity. This property introduces new trade-

off aspect for HOS signal processing. The additional variance introduced by PCS-HOS is able to be suppressed by extending the length of signal, adding preprocessing filters, or using overlapped window.

In the simulation, we also observe that PCS-HOS(4) is not suitable for estimating cumulants. But concerning the advantages of fourth-order HOS over its third-order counterpart, it is worthwhile to find out new approaches suppressing the variance caused by both PCS and HOS. There are also other aspects worth further exploring, such as the influence of filter in preprocessing, choices of large PCS, and implementation of PCS with nonparametric methods for channel estimation.

CHAPTER 5

Nested sampling for HOS with application to channel estimation

5.1 Introduction

Accurate channel estimation can be used in orthogonal frequency division multiplexing systems to improve the performance by allowing coherent demodulation, which possesses a 3-to-4 dB gain in signal-to-noise ratio (SNR) compared with differential modulation [64]. Besides, systems are able to take advantage of receiver diversity for optimum combining via channel estimators. In general, channel estimation can be conducted by either inserting pilot segments into OFDM subcarriers [65], or estimating channel only based on the received signal. But it is not efficient for repeated transmitting a known sequence to train the equalizer at the receiver, and it is difficult to establish data transmission over the channels suffered from unavoidable presence of multipath fading. Theoretically speaking, the blind channel estimation is preferable.

There are two kinds of approaches to implement the blind channel estimation. The linear methods are based on the fact that the cyclostationary characteristics of modulated signals permit the recovery of both amplitude and phase responses of a communication channel only using second-order statistics [66]. The nonlinear methods are based on the calculation of higher-order statistics (HOS), and could be further divided into direct or indirect approaches. The direct algorithm calculate HOS and their discrete Fourier transforms with further matrix manipulation [67]. Although its computation is more complex, the advantages include more straightforward for theoretical derivation, free from minimizing cost functions so as to avoid local mini-

mum problem, and insensitivity to time jitter. On the other hand, one of the typical indirect algorithms is Bussgang algorithm [68], in which the deconvoluted signal implicitly exploit the HOS via Bussgang statistics. This kind of methods is simpler to implement and generally capable of delivering a good performance.

In general, HOS approaches are favorable to preserve both phase and amplitude information of the signal, and to deal with non-Gaussian sequences and nonlinear non-minimum phase system. Its estimates can be calculated from either conventional nonparametric methods of Fourier transform [23], or parametric methods based on moving average (MA), autoregressive (AR), or autoregressive moving average (ARMA) models [69]. The paper [23] provides further details about both kins of methods. It is worth noting that both of them have high variance and require a large number of records to obtain smooth estimates [24], but increasing the number of segments is demanding on computation, and may increase bias and introduce nonstationarity. Although there are preliminary researches [69, 70] providing a sparse sampling scheme, pairwise coprime sequences (PCS), as a possible solution, how to accelerate speed of convergence and to leverage the trade-off between complexity and performance are still open questions.

This is our motivation to design an algorithm estimating HOS based on nested sampling. This concept is first introduced to conduct economic and sociological surveys [71]. Later researches extend it to array signal processing [6], astronomical data analysis [72], and estimating the directions of arrivals [7, 73, 74]. But nested sampling has never been considered to derive HOS in previous literatures.

The main purpose of this paper is to develop a HOS algorithm based on multilevel nested sampling (MNS). The direct merits of implementing nested sampling are to greatly reduce the computational expense and to produce posterior inferences at the same time. By using this sampling scheme, the proposed algorithm improves the

PCS performances in terms of estimation variance and speed of convergence with the same delay and a even lower computational complexity.

In Section 5.2, we discuss theoretical background about both higher-order statistics and nested sampling. Then, we describe the MNS-HOS algorithm, and analyze its complexity and complexity in Section 5.3. The performance of the proposed algorithm is simulated in Section 5.4. Finally, we draw the conclusions in Section 5.5.

5.2 Preliminaries for nested sampling

Nested sampling is a non-uniform sampling scheme, using two different samplers in each of given period. Given sampling interval T_s, although the signal is sparsely and non-uniformly sampled at $1 \leq l \leq N_1 T_s$ and $(N_1+1)mT_s$, for $1 \leq m \leq N_2$, in one sampling period, the autocorrelation $c_{2x}(\tau)$ is able to be estimated at all lags $\tau = kT_s$, where k, l, and m are integers.

In the simplest form, the nested array has two levels of sampling density, with the level 1 samples at the N_1 locations and the level 2 samples at the N_2 locations:

$$\mathscr{L}_{inner} = \{mT_s,\ m = 1, 2, \ldots, N_1\},$$
$$\mathscr{L}_{outer} = \{n(N_1+1)T_s,\ n = 1, 2, \ldots, N_2\}.$$

As a result, the sampling interval for the inner layer is $t_1 = T_s$, while the outer layer is $t_2 = (N_1+1)T_s$.

Figure 5.1 shows an example of periodic sparse sampling using nested sampling structure with N_1=3 and N_2=5. The cross-difference between two levels is

$$k = (N_1+1)m - l, \quad 1 \leq m \leq N_2,\ 1 \leq l \leq N_1, \tag{5.1}$$

and its range is given by

$$k \in [-(N_1+1)N_2 - 1, (N_1+1)N_2 - 1]\,. \tag{5.2}$$

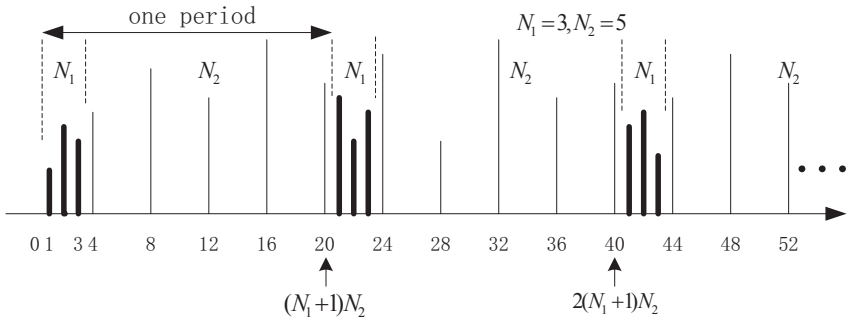

Figure 5.1: Nested Sampling with N_1=3, N_2=5

Note that there are missing values symmetric through zero within the range in (5.2). The positive part includes $(N_1+1), 2(N_1+1), \ldots, (N_2-1)(N_1+1)$. Take the Figure 5.1 as an example, the cross-difference will achieve the indexes

$$1, 2, 3, (), 5, 6, 7, (), 9, 10, 11, (), 13, 14, 15,$$

where the values with indexes in parentheses are left unknown. Then, we can use the self-difference among points from the second level to calculate for the missing indexes, which are 4, 8, and 12 in this example. The calculation is generalized as

$$(N_1+1)(m_1-m_2), \quad 1 \leq m_1, m_2 \leq N_2. \tag{5.3}$$

With this method, we are able to calculate the second-order statistics without the constraint about the lowest permitted sampling rate.

5.3 Multilayer nested sampling HOS algorithm

5.3.1 MNS-HOS algorithm

In this Section, we propose the MNS-HOS algorithm and make the nested sampling according to the instantaneous variance of the input signal for better HOS estimation. Besides, we also demonstrate how this algorithm reduces the computational complexity by several orders. In the following content, MNS $y_0(n), y_1(n), \ldots, y_k(n)$

denote the sampled sequences from output signal related with the nested level $N_0, N_1, \ldots, N_k$, respectively. In addition, we have the following assumptions to modeling the system:

S.1 The driving signal $\{w(t)\}$ is non-Gaussian, zero-mean, and independent and identically distributed (i.i.d.). Besides, it only has finite moments, which are estimated via expectations, and denoted as $E[w^2(t)] = \sigma_w^2$, $E[w^3(t)] = \gamma_{3w}$, and $E[w^6(t)] < +\infty$.

S.2 The measurement noise $\{v(t)\}$ is assumed to be zero-mean, and to have either white or color Gaussian distribution. It is also i.i.d. and independent of $\{w(t)\}$ with $E[v^2(t)] = \sigma_v^2$.

S.3 The channel is regarded as a linear, time-invariant, and non-minimum phase moving average (MA) system, and its order $q+1$ has known beforehand.

Based on these assumptions, we derive a procedure for estimating the covariances of the sampled cumulants, and model the output of the channel and the received signal as

$$x(n) = \sum_{i=0}^{q} h_i w(n-i), \tag{5.4}$$

$$y(n) = x(n) + v(n). \tag{5.5}$$

Then, we have

$$\begin{aligned}
&E[\hat{c}_{ky}(s_1, \ldots, s_k), \hat{c}_{ky}(t_1, \ldots, t_k)] \\
&= E\left[\frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \prod_{i=0}^{k-1} y_i(m+s_i) \prod_{j=0}^{k-1} y_j(n+t_j)\right] \\
&= E\left[\frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \prod_{i=0}^{k-1} y_i(s_i) \prod_{j=0}^{k-1} y_j(n-m+t_j)\right] \\
&= \frac{1}{N} \sum_{n=-(N-1)}^{N-1} \left(1 - \frac{|n|}{N}\right) E\left[\prod_{i=0}^{k-1} y_i(s_i) \prod_{j=0}^{k-1} y_j(n+t_j)\right],
\end{aligned} \tag{5.6}$$

where $s_0 = t_0 = 0$. Hence,

$$\begin{aligned}
&\operatorname{cov}[\hat{c}_{ky}(s_1,\ldots,s_k),\hat{c}_{ky}(t_1,\ldots,t_k)] \\
=&\frac{1}{N}\sum_{n=-(N-1)}^{N-1}\left(1-\frac{|n|}{N}\right)E\left[G(y,k,0,s)G(y,k,n,t)\right]. \qquad (5.7)
\end{aligned}$$

Since $\{y(n)\}$ is a MA(q) process, the product from $y(0)$ to $y(s_k)$ is statistically independent of the product form $y(n)$ to $y(n+t_k)$ for $n > s_1 + q$ and $n < -(t_1 + q)$. Therefore, the (5.7) can be written as

$$\begin{aligned}
&\operatorname{cov}[\hat{c}_{ky}(s_1,\ldots,s_k),\hat{c}_{ky}(t_1,\ldots,t_k)] \\
=&\frac{1}{N}\sum_{n=-(t_1+q)}^{s_1+q}\left(1-\frac{|n|}{N}\right)E\left[G(y,k,0,s)G(y,k,n,t)\right] \\
\approx&\frac{1}{N^2}\sum_{n=-(t_1+q)}^{s_1+q}\left(1-\frac{|n|}{N}\right)\hat{G}(y,k,0,s)\hat{G}(y,k,n,t), \qquad (5.8)
\end{aligned}$$

where

$$\begin{aligned}
G(y,k,n,t) &= \prod_{j=0}^{k-1} y_j(n+t_j) - c_{ky}(t_1,\ldots,t_k) \\
\hat{G}(y,k,n,t) &= \prod_{j=0}^{k-1} y_j(n+t_j) - \hat{c}_{ky}(t_1,\ldots,t_k).
\end{aligned}$$

The approximation in the second step of (5.8) is based on the stationary and ergodicity properties of the given process, which provides the consistent estimation of the expectation in its first step.

Specifically, for the third-order cumulant, we can derive the variance of estimation based on its covariance, and let $\tau_1 = s_1 = t_1$ and $\tau_2 = s_2 = t_2$, we have the variance of sampled third-order cumulants

$$\begin{aligned}\hat{\sigma}^2[\hat{c}_{3y}(m,n)] \approx \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=-q+\tau_2}^{q+\tau_1}\left(1-\frac{|j|}{N}\right) \\ \cdot [y_0(n)y_1(n+\tau_1)y_2(n+\tau_2) - \hat{c}_{3y}(\tau_1,\tau_2)] \\ \cdot [y_0(m+n)y_1(m+n+\tau_1)y_2(m+n+\tau_2) - \hat{c}_{3y}(\tau_1,\tau_2)]. \end{aligned} \tag{5.9}$$

The multi-level nested sampling (MNS) is derived from the two-levels nested sampling, and can be easily related with the calculation of HOS.

$$\begin{aligned}\mathscr{L}_1 &= \{nT_s, \quad n = 1, 2, \ldots, N_1\}, \\ \mathscr{L}_i &= \left\{nT_s\prod_{j=1}^{i-1}(N_j+1),\ n = 1, 2, \ldots, N_i\right\},\ i = 2, \ldots, K,\end{aligned}$$

where K is the number of levels, and $\{N_i\}_{i=1}^{K}$ are number of samples for the ith layer.

Compared with the basic coprime sampling and the pair-wise coprime sampling for HOS [69], the MNS has no constraint on the choice of downsampling factor N_i. Furthermore, given the sample points are independent and i.i.d., the sampling scheme is able to take advantage of less noisy segments for better estimation.

Based on the analysis of variance, we can summarize the MNS-HOS algorithm as follow:

5.3.2 Complexity analysis

The following content shows that the complexity improvement is nontrivial via estimating the HOS based on MNS. To quantify the complexity, we use the "Big O" notation to denote the asymptotic limitation of a function when the argument increases towards infinity. Both the parametric and nonparametric method have to

Algorithm 1 MNS-HOS Algorithm

Require: $N_i > 0, \forall i \in [0, K]$, $\{y(n)\}_n^N$, for $N \geq \prod_{i=0}^{K-1} N_i$

Initialize signal average $\hat{\mu}$ based on previous segment

for $i = 1$ to N **do**

Update average $\hat{\mu}$

Calculate variance for every segment with length $N_1 + 1$.

if $i \equiv 0 \mod N_j,\ \forall j = 2, \ldots, k-1$ **then**

Average N_j variances with interval $\prod_{l=1}^{j-1} N_l$

end if

end for

for $i = k$ to 1 **do**

Nested range $\leftarrow$ range with the smallest variance

end for

Conduct multilevel nested calculation

estimate cumulant based on expectation, which is the first part of the complexity. It mainly comes from multiplication traversal in (3.1), (3.2), and (3.3). The operations must be performed record by record which in turn causes the complexity to increase exponentially. To estimate kth-order cumulant, the computation complexity is $O(N^k)$ where N is the length of segment.

The nonparametric method further requires frequency averaging which makes it more impractical for longer sequences. The classical FFT algorithm achieves computational complexity $O(NlogN)$ for second-order spectrum and $O(N^{r-1}logN)$ for the rth-order Fourier transform [31]. Including the calculation of cumulants, the overall complexity achieves as high as $O(N^{2k-1}logN)$.

On the other hand, MNS is able to decrease the calculation of cumulant by at least one order. Suppose the length of segment is $2N$. The complexity to calculate the third-order cumulant is

$$\frac{2N}{N_1} \times \frac{2N}{N_2} \times \frac{2N}{N_3} \sim O(N^2), \tag{5.10}$$

where $N = \prod_{i=1}^{3} N_i$. Note that the variance estimation produce the overhead $O(N)$ to the calculation, but it is trivial compared with the HOS procedure and can be ignored in the complexity analysis.

For nonparametric approach, the optimization is more significant by making the Fourier transform in-place without complicated index mapping. The following analysis is inspired by the structure of prime-factor FFT in [33]. For the kth-order spectrum, we consider (3.4) to be a series of two-dimensional Fourier transform. Then, the formula is rewritten as

$$\begin{aligned} &S_{kx}(\omega_1, \omega_2, \ldots, \omega_{k-1}) = \\ &\frac{1}{L} \sum_{l=1}^{L} \sum_{\tau_{k-1}=-N}^{N} \left[\ldots \left[\sum_{\tau_1=-N}^{N} c_{kx}(\tau_1, \tau_2, \ldots, \tau_{k-1}) \right.\right. \\ &\left.\left. exp\left(\frac{-j2\pi\omega_1\tau_1}{N}\right)\right] \ldots \right] exp\left(\frac{-j2\pi\omega_{k-1}\tau_{k-1}}{N}\right). \end{aligned} \tag{5.11}$$

Specifically, we can simplify the bispectrum as

$$\begin{aligned} &S_{3x}(\omega_1, \omega_2) = \sum_{(a_1-a_3)=-N}^{N} \left[\sum_{(a_1-a_2)=-N}^{N} \right. \\ &x_1(a_1) exp\left(\frac{-j2\pi a_1(\omega_1+\omega_2)}{N}\right) \\ &\left. x_2(a_2) exp\left(\frac{-j2\pi a_2\omega_1}{N}\right)\right] x_3(a_3) exp\left(\frac{-j2\pi a_3\omega_2}{N}\right). \end{aligned} \tag{5.12}$$

The averaging among L segments is deliberately omitted for simpler formula abstraction. It can be easily restored in the implementation. Because L increases

linearly along with the length of signal, it is negligible compared with the exponentially increased asymptotic property of N. In (5.11), complex multiplications are in-place calculated, which only change a constant factor of the asymptotic limitation. Hence, the calculation procedure decreased from $O(N^5 logN)$ of the existing HOS to quadratic form $O(N^2)$ for MNS-HOS.

The fundamental difference between (5.12) and (3.4) is that without MNS, c_{kx} is calculated by the exhausted multiplying among N records for k times, which cannot be decomposed.

It is also worth mentioning that the MNS-HOS algorithm has the same order of complexity as another sparse sampling scheme PCS-HOS [69]. Compared with PCS-HOS, MNS-HOS uses an iteration for estimating the variance of the signal, keeps updating the sample average, and requires an extra memory with the length of window size to record the variances. As a result, MNS-HOS is able to better explore the statistical characteristics of the signal. The detailed improvement is provided in the following section.

5.4 Simulation results

5.4.1 Basic setting

In order to fully explore properties of MNS, we apply it to estimate the simplified LTE spacial channel model [34] where the channel is assumed to be a MA model with six paths. Furthermore, we simplify each path to be only one sub-path, and assume the channel characteristics remain invariant over the processed data symbols, which

means the channel is slow fading for the very high signaling rates. MA(2) and MA(5) are simulated according to the system model in (5.4) and (5.5):

$$x_1(n) = w(n) - 2.333w(n-1) + 0.667w(n-2) \tag{5.13}$$

$$\begin{aligned} x_2(n) =& w(n) + 0.1w(n-1) - 1.87w(n-2) \\ &+ 3.02w(n-3) - 1.435w(n-4) \\ &+ 0.49w(n-5) \end{aligned} \tag{5.14}$$

$$y_1(n) = x_1(n) + v_1(n) \tag{5.15}$$

$$y_2(n) = x_2(n) + v_2(n). \tag{5.16}$$

In both models, we assume $v_1(n)$ and $v_2(n)$ are additive white Gaussian noise (AWGN). The driven signal $w(n)$ is a zero-mean exponential random deviate process with $E[w^2(n)] = 1$, $E[w^3(n)] = 2$, and $E[w^4(n)] = 9$. We generate signal sequence with 700000 sample points for a single run and average the results from 50 times of Monte Carlo simulations for each setting. There is no filter used for preprocessing the input signal, and the overlap percentage is set to be zero. We select the coefficients N_1=4, N_2=3, and N_3=4 for MNS-HOS, which makes it decrease the complexity to 2.1% of the original HOS algorithm. As comparison, the PCS-HOS uses N_1=2, N_2=3, and N_3=5, which decreases the complexity to 3.3%.

Because the absolute value of MA coefficients vary from 0.1 to 3.02, the variance itself cannot fully reveal how well the estimation is. We use the ratio between standard deviation and its expectation as vertical axis to indicate the convergence.

5.4.2 Estimating cummulants via MNS and PCS

Figures 5.2 shows the comparison between third-order MNS and PCS for estimating cumulants with 2 taps. The ratio between the standard deviation of MNS and

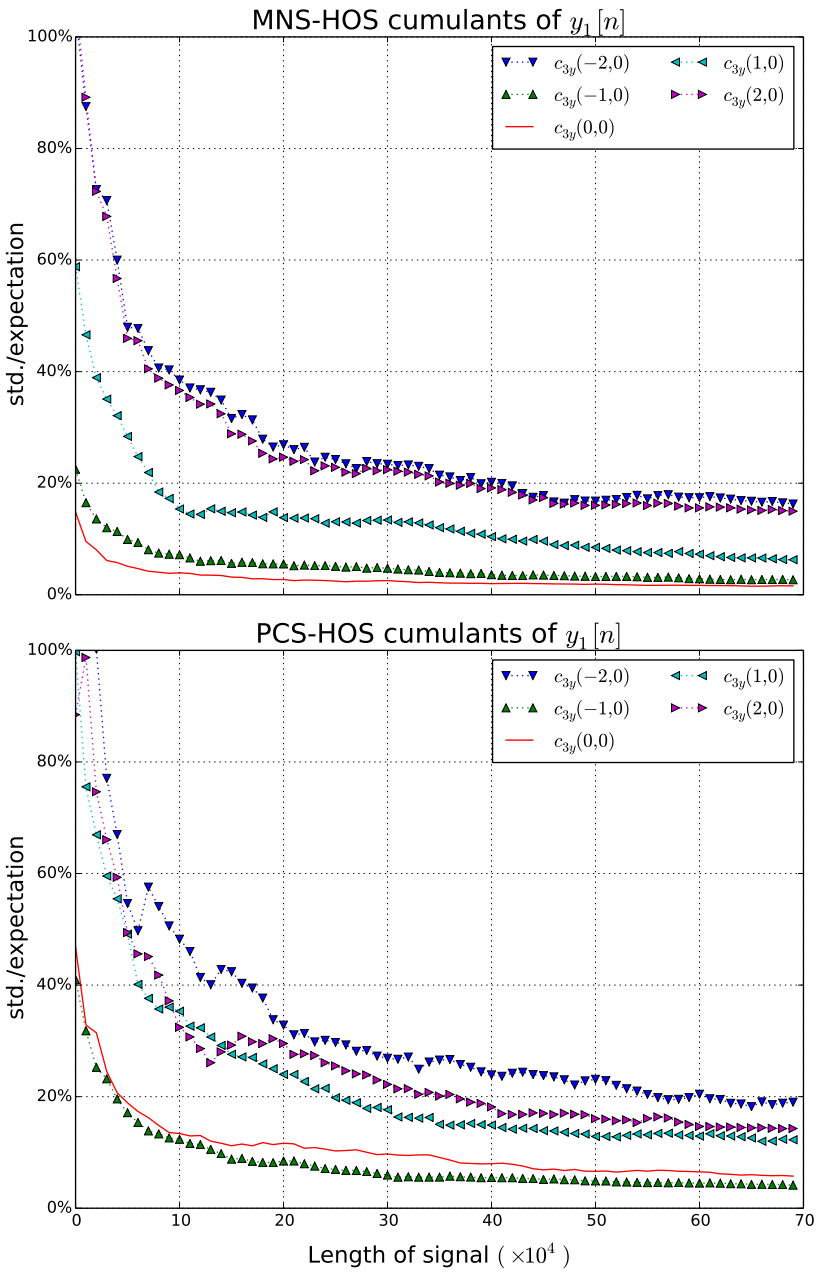

Figure 5.2: Comparison of cumulant convergence for $y_1[n]$

its expectation falls below 40% of the expected value after $100K$ samples, and below 20% after $400k$ samples, while the ratio of PCS falls below 40% at after receiving $170K$ samples, and the $c_{3y}(2,0)$ falls below 20% after $600K$ samples.

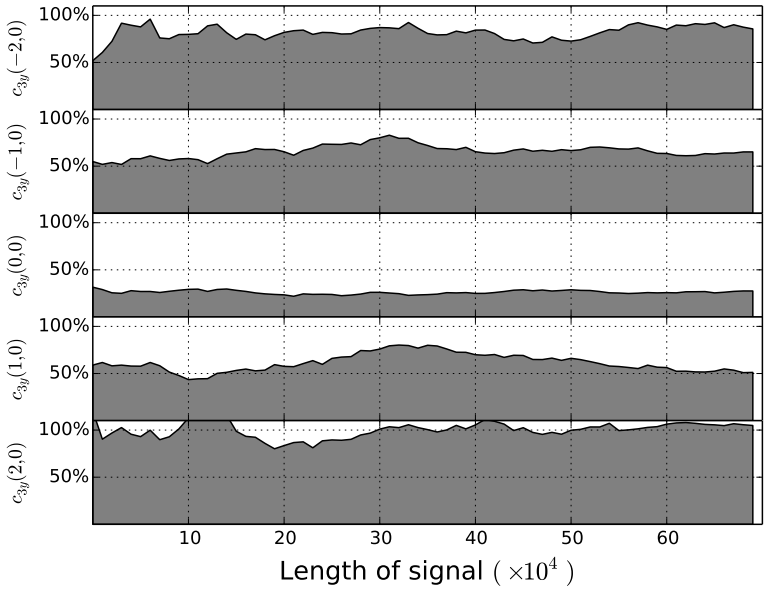

Figure 5.3: Ratios of variances between MNS and PCS

Figure 5.3 visualizes the comparison of each taps, in which the percentages in the y axis stand for the ratios between MNS and PCS from Figure 5.2, and the tap number is labeled at the left side of the figure. We can see that the most obvious improvement of MNS upon PCS is for the 0 and ± 1 taps.

Figures 5.4 indicates the convergence of both algorithm for MA(5) model. Compared with Figures 5.2, the estimates of both algorithms suffer from increased variance and decreased speed of convergence. However, MNS also performs much better than PCS. As fitting both curves into the same scale for y axis, most of the MNS ratios are overlapped within 100% area, while PCS has a much scattered curves indicating larger variance and more inferior convergence.

5.4.3 Estimating MA system via MNS and PCS

In Figure 5.5, the performances of HOS algorithms for MA system identification [35] are quantified via root-mean square error (RMSE). Three candidates in the

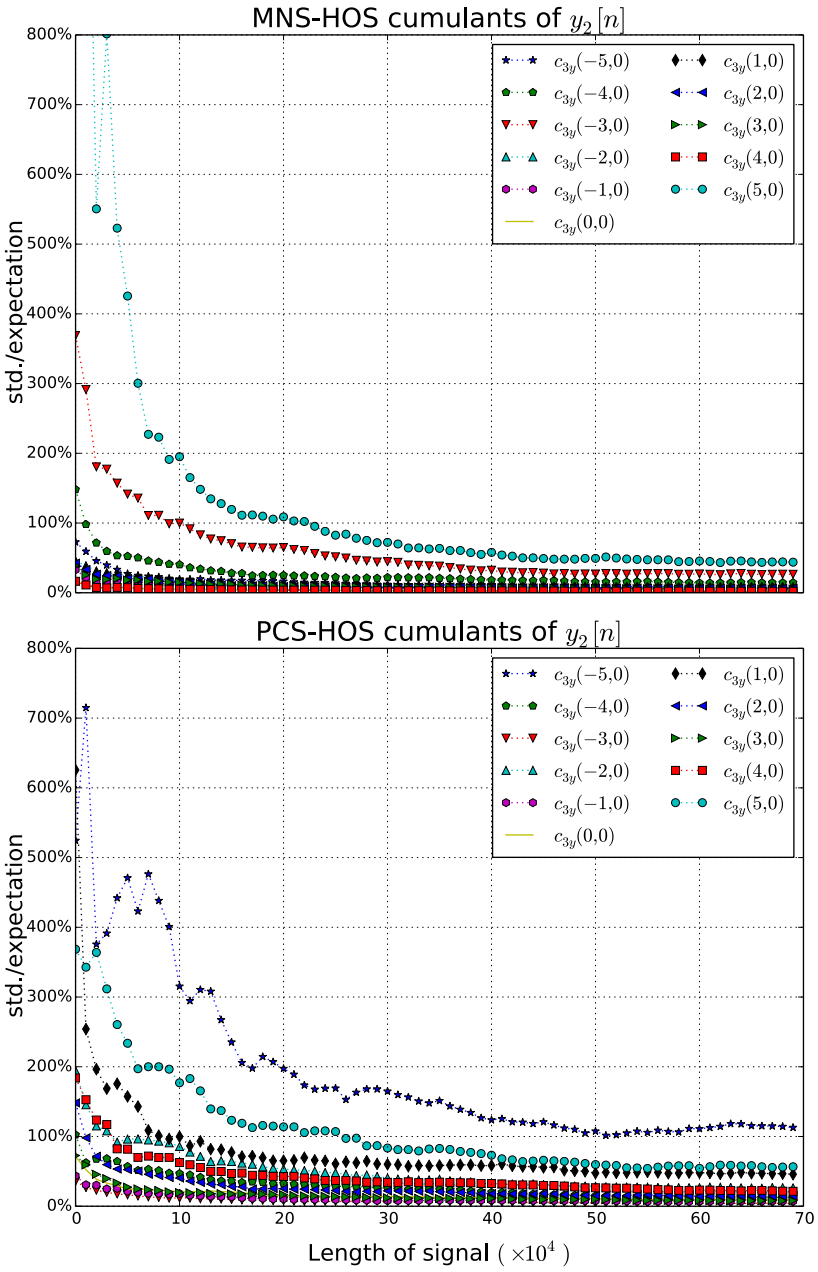

Figure 5.4: Comparison of cumulant convergence for $y_2[n]$

comparison are grouped by the RMSE for each tap, including PCS-HOS in the left side (blue), MNS-HOS in the middle (green), and existing HOS in the right side (red).

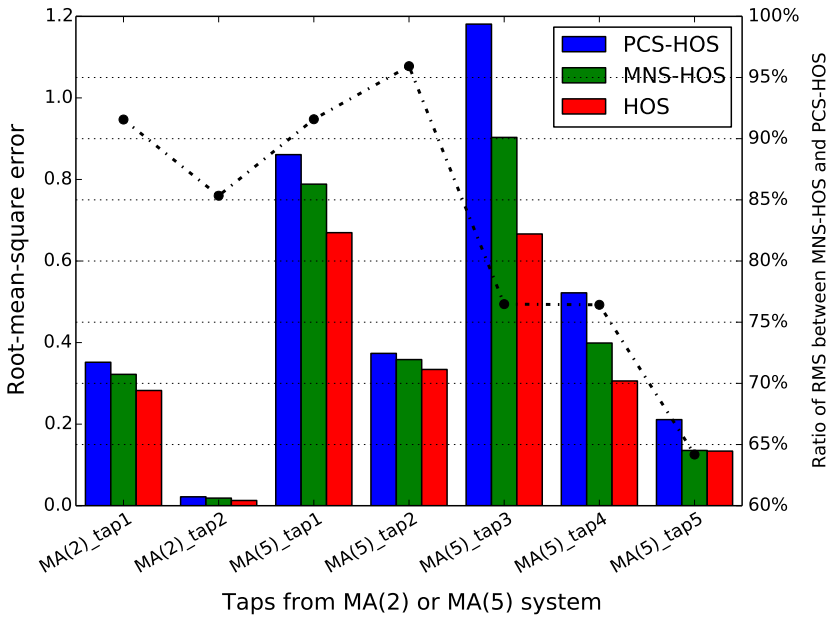

Figure 5.5: Performance comparison

The left vertical axis indicates the values of RMSE. The order of performances from best to worst is HOS, MNS-HOS, and PCS-HOS.

The broken line using the right vertical axis indicates the performance improvement of MNS-HOS relative to PCS-HOS. Generally speaking, MNS-PCS consistently performs better than PCS-HOS. For MA(2) model, the improvement is not significant because all the algorithms have a relative low RMSE in short MA systems. The performance gain becomes remarkable in MA(5) model. MNS-HOS outperforms PCS-HOS by 16.9% less RMSE on average. Specifically, it has 17.9% less RMSE in MA(2) model and 8.8% less RMSE in MA(5) model. It is also desireable to observe that the last three taps of MA(5) system do not deteriorate the estimates of MNS-HOS. It can be explained by the good convergence of $c_{3y}(\pm 5, 0)$ and $c_{3y}(\pm 4, 0)$ in the upper figure of Figure 5.4, because large variation for the cumulants of longer delay is the major factor undermining the performance of corresponding MA coefficients. Concerning

the MNS-HOS in this comparison has lower complexity (2.1% of the HOS without sparse sampling) than the PCS-HOS has (3.3%), it is favorable to involve MNS-HOS in the calculation of HOS when we leverage the trade-off between performance and complexity.

Nevertheless, it is worth mentioning that the performance gain and complexity simplification are acquired at the expense of increasing the delay of the system to average longer sequence, which is the case for both MNS-HOS and PCS-HOS algorithms. For the same length of signal sequences, existing HOS has about fifty times more computation. MNS-HOS has about 21.6% performance loss on average–generating 15.4% more RMSE in MA(2) and 22.5% more in MA(5).

5.5 Conclusion

In order to use the nested sampling to calculate HOS, we use multilevel nested sampling to develop the MNS-HOS algorithm. This algorithm take advantage of the second-order statistics during calculating HOS via sparse sampling, which significantly reduces the computational demands for calculating HOS while still maintains the statistical properties.

In the simulation, MNS-HOS is first compared with PCS-HOS algorithm. It produces less variance and converges faster for estimating HOS cumulants. As the taps of MA prolong from two to five, these advantages become more obvious. We further apply it to channel estimation where MA models are assumed, and compare its performance with both original HOS algorithm and PCS-HOS. The MNS-HOS is able to achieve 16.9% performance gain relative to PCS-HOS, and has 21.6% loss of performance as opposed to the HOS without sparse sampling, but only uses its 2.1% complexity cost.

Although MNS-HOS is superior in the context of sparse sampling, we observe that it still converges much slower than the original HOS does. It only handles the problem of computational complexity in HOS calculation. Further improvement could be attained by introducing Newton-Raphson method or Gauss-Newton algorithm. Besides, this paper only discusses the third-order HOS. Extending the algorithm to even higher moments is meaningful and nontrivial. There are also other aspects worth further exploring, such as the influence of filter in preprocessing, adaptively changing the choices of nested factors, and implementation of MNS with nonparametric methods for channel estimation.

CHAPTER 6

Capacity improvement of multi-cell cooperative cellular networks with nested deployment

6.1 Introduction

Cellular networks are important parts of wireless communication systems. As the number of macro-cellular BSs reaching 50 million worldwide by 2015 [75], it has driven intensive research to fully utilized these facilities catering the growing demand for high-data rate services.

There are quite a few works [76] indicating that sophisticated cooperations among multiple BSs could achieve enormous gain over the lone-BS model. By increasing the level of collaboration, the multiple BSs are regarded as a distributed antenna array, and the multi-input multi-output (MIMO) techniques can be used to increase capacity, decrease interference, and perform distributed beamforming. Specifically, the downlink channel is commonly considered as a Gaussian noise plus interference broadcast channel (BC), and applied single-user detection because of the complexity and power constraints of the mobile receivers. The uplink channel is usually modeled as a multiple-access channel (MAC), because there are less restrictive limitations at the BSs, and the received signals are joint processed for the system enhancement [77].

For the uplink channel, an early study by Wyner [78] presented an analytically tractable model, in which the cells were ordered in an infinite linear vector or 2D hexagonal matrix. It derived optimal throughput and linear minimum mean-square error for non-fading channels. These results were extended to flat-fading channels

in [79], where the fading was observed to increase the throughput under certain circumstances.

For the downlink channel, the sum-rate of a multiple-antenna cellular system was addressed with power constraints for each BS in [80] by taking advantage of the duality principle between the BC and MAC [81]. "Dirty paper coding" (DPC) principle [82] was also applied to eliminate the effect of uncorrelated additive interference. The corresponding achievable rate and improvements of spectrum efficiency and sum-rate capacity for fading channel were reported in [83–85].

However, on the other hand, there are several inherent problems. First of all, because of the limitation of backhaul links, it is impossible to recruit an arbitrary number of BSs in order to achieve enormous spectrum efficiency gains [86]. Besides, if the cooperative network is modeled as a cluster and put into a larger system that one cluster is adjacent to the others, the out-of-cluster interferences would hinder further improvement as the per-cluster transmit power increases [76].

In this paper, a novel nested-distributed network is introduced to advance the research on joint multi-cell processing. It is also the first literature analyzing the cellular system with the nested-distributed BSs from the information-theoretic point of view. The nested array is firstly introduced to perform array processing with increased degrees of freedom using much fewer physical sensors [6]. The paper [73] generalizes this concept to the multiple dimensions, and provides the optimal structure to maximize the number of elements in the virtual co-array, as well as derives closed-form expressions for the sensor locations and the exact degrees of freedom obtainable from the proposed array as a function of the total number of the sensors.

The paper is organized as follows. Section 6.2 introduces the theoretical backgrounds for both 2D nested array and sum-rate capacity of cooperative multi-cell processing, as well as the basic assumptions for the network model. Section 6.3 de-

rives the invariance of the difference co-array in order to obtain the capacity of nested distributed network. In Section 6.4, the numeric results are provided to validate the proposed propositions. Finally, we summarize the paper in Section 6.5, and provide the directions for further research.

6.2 Preliminary and model description

6.2.1 2D nested co-array

The first two concepts relate with multidimensional lattice, which is extensively used to represent the nest co-array.

Definition 3 (Fundamental Parallelepiped (FPD) [73]). *The FPD of* $\boldsymbol{V} \in \mathcal{C}^{D\times D}$ *in D dimensions FPD(*$\boldsymbol{V}$*) is defined as the set of all vectors of the form*

$$\{\boldsymbol{V}\boldsymbol{x}, \boldsymbol{x} \in [0,1)^D\}.$$

Visually, *FPD*($\boldsymbol{V}$) consist of all points contained in the parallelepiped whose sides are given by the two column vectors of $\boldsymbol{V}$. Its volume is given by $|\det(\boldsymbol{V})|$ and the density of the lattice points is the inverse of its volume.

Definition 4 (Shifted FDP (SFPD)). *For arbitrary integers* k_1 *and* k_2*, the SFPD is defined as the FPD(*$\boldsymbol{N}^{(s)}$*) shifted by the vector* $[k_1, k_2]^T$*, which is*

$$SFPD(\boldsymbol{N}^{(s)}, k_1, k_2) \triangleq \left\{\boldsymbol{N}^{(s)}([k_1, k_2]^T - x), x \in [0,1)^2\right\}.$$

The following definition provides one of the configuration for nested co-array deployments, and the consecutive theorem guarantees that this Definition 5 satisfies the requirements of both FPD and SFPD.

Definition 5 (2D Nested Co-Array). *A two dimensional nested array is described by a* 2×2 *non-singular matrix* $\boldsymbol{N}^{(d)}$*, an integer matrix* $\boldsymbol{P}$ *and integers* $N^{(s)}$*,* $N^{(d)} = det(\boldsymbol{P})$*, and satisfying*

1. *A dense array with* $N^{(d)} = det(\boldsymbol{P})$ *elements on lattice generated by* $\boldsymbol{N}^{(d)}$*, with sensor locations given by* $\{\boldsymbol{N}^{(d)}\boldsymbol{n}^{(d)}, \boldsymbol{n}^{(d)} \in FPD(\boldsymbol{P})\}$.
2. *A sparse array* $\boldsymbol{N}^{(s)} = \boldsymbol{N}^{(d)}\boldsymbol{P}$*, with sensor locations given by* $\boldsymbol{N}^{(s)}[k_1, k_2]^T$ *with* $0 \leq k_1 \leq N_1^{(s)} - 1$, $0 \leq k_2 \leq N_2^{(s)} - 1$.
3. $N_1^{(d)} N_2^{(d)} = N^{(d)}$ *and* $N_1^{(s)} N_2^{(s)} = N^{(s)}$*, where* $N_1^{(d)}$ *and* $N_2^{(d)}$ *are the number of rows and columns for the dense array, and* $N_1^{(s)}$ *and* $N_2^{(s)}$ *are corresponding numbers for the sparse array.*

Theorem 4. *Consider two nonsingular* 2×2 *matrics* $\boldsymbol{N}^{(s)}$ *and* $\boldsymbol{N}^{(d)}$ *related by an integer matrix* $\boldsymbol{P}$ *as* $\boldsymbol{N}^{(s)} = \boldsymbol{N}^{(d)}\boldsymbol{P}$.

1. *Any point on* $\boldsymbol{N}^{(d)}\boldsymbol{n}$ *on the lattice* $\boldsymbol{N}^{(d)}$ *can be expressed as* $\boldsymbol{N}^{(d)}\boldsymbol{n} = \boldsymbol{N}^{(s)}\boldsymbol{n}^{(s)} - \boldsymbol{N}^{(d)}\boldsymbol{n}^{(d)}$ *where* $\boldsymbol{n}^{(s)}$ *is an integer vector and* $\boldsymbol{n}^{(d)} \in FPD(\boldsymbol{P})$.
2. *All points within* $SFPD(\boldsymbol{N}^{(s)}, k_1, k_2)$ *can be generated by the differences* $\{\boldsymbol{N}^{(s)}[k_1, k_2]^T - \boldsymbol{N}^{(d)}\boldsymbol{n}^{(d)}, \boldsymbol{n}^{(d)} \in FPD(\boldsymbol{P})\}$.

6.2.2 Sum-rate capacity for multi-cell processing

The ergodic per-cell sum-rate capacity is given by [78]

$$C(P) = \frac{1}{L} E\left[\log_2\left(\boldsymbol{I}_L + P\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right)\right] \tag{6.1}$$

where P is the transmit power of a single user, and the expectation is taken with respect to the fading coefficients $\boldsymbol{H}_L$. The matrix $\boldsymbol{H}_L\boldsymbol{H}_L^\dagger$ is an $L \times L$ matrix given by

$$\left[\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,n} = \begin{cases} \boldsymbol{a}_m\boldsymbol{a}_m^\dagger + \boldsymbol{b}_m\boldsymbol{b}_m^\dagger & m = n \\ \boldsymbol{b}_m\boldsymbol{a}_n^\dagger & n = (m-1) \bmod L \\ \boldsymbol{a}_m\boldsymbol{b}_n^\dagger & n = (m+1) \bmod L \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

The definition of spectrum efficiency, denoted as γ, is expressed as a function of the system average transmit E_b/N_0 [89]. Its value is solved by substituting

$$P = \frac{E_b}{KN_0}\gamma \tag{6.3}$$

in (6.1), and K is the number of users in a cell. The $\frac{E_b}{N_0}$ required for reliable communication is defined as

$$\frac{E_b}{N_{0\,min}} \triangleq \frac{log_e 2}{C'(0)} \tag{6.4}$$

Then, the slope of the spectrum efficiency for low-SNR and high-SNR can be defined as

$$\begin{aligned} S_{low} &\triangleq -\frac{2\left[C'(0)\right]^2}{C''(0)} \\ &= \lim_{\frac{E_b}{N_0}\to\frac{E_b}{N_{0\,min}}} \frac{10\gamma\log_{10}2}{10\log_{10}\frac{E_b}{N_0} - 10\log_{10}\frac{E_b}{N_{0\,min}}} \end{aligned} \tag{6.5}$$

$$S_{high} \triangleq \lim_{P\to\infty} PC'(P) = \lim_{\frac{E_b}{N_0}\to\infty} \frac{10\gamma\log_{10}2}{10\log_{10}\frac{E_b}{N_0}} \tag{6.6}$$

The upper and lower bounds for the fading channels are

Theorem 5. *For $K \gg 1$, the average per-cell sum-rate capacity for Rayleigh fading satisfies*

$$\begin{aligned} &\log_{10}(1 + KP((1-\varepsilon)\log_e K + 2)) \\ &\qquad \le C(P) \le \log_{10}(1 + 2KP\log_e K). \end{aligned} \tag{6.7}$$

Theorem 6. *The spectrum efficiency of channel for Rayleigh fading are characterized, for any number of BSs $M \ge 3$, by*

$$\begin{aligned} &S_{low} = 2; \quad S_{high} = 1; \\ &\frac{(\log_e 2)^2}{2\log_e K} \le \frac{E_b}{N_{0\,min}} \le \frac{(\log_e 2)^2}{(1-\varepsilon)\log_e K + 2}. \end{aligned} \tag{6.8}$$

In both theorems, the $\varepsilon = Pr\{P > P_{out}\}$, indicating the possibilities that certain number of users passes a lower bound of the fading power that is needed to transmit. For example, the possibility that at least one of the users at a give cell cross the threshold is $1-(1-1/K^{(1-\varepsilon)})^K$. If $K = 100$ and $\varepsilon = 0.1$, the possibility is 79.762%, and 99.997% for $\varepsilon = 0.5$. Refer to [85] for detail proofs of these two theorems.

6.2.3 Description of the system model

Figure 6.1 illustrates an example deploying BSs in a 2D nested array manner. Every hexagon stands for a macrocell and the colored ones are deployed with actual BSs. It could be regarded as a simplified model of metroplex area containing both urban and suburb regions. Specifically, the sparse array is the lattice with blue color indexing $(x, 1)$, where $x \in [1, 14]$, and the dense array is the yellow hexagons in the middle.

As a result, its difference co-array is $\boldsymbol{N}^{(d)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{P} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$, $N_1^{(s)} = 3$ and $N_2^{(s)} = 3$. Both the lattices towards upper left and lower right are regarded as the positive halves of the difference co-array, whose elements are given by

$$\left\{ \boldsymbol{N}^{(d)} [k_1, k_2]^T, \; -(p_1 - 1) \le k_1 \le \left(N_1^{(s)} - 1 \right) p_1, \right. \\ \left. -(p_2 - 1) \le k_2 \le \left(N_2^{(s)} - 1 \right) p_2 \right\}.$$

where p_1 and p_2 are the eigenvalues of the $\boldsymbol{P}$.

Based on the Theorem 4, we can conclude that the positive half of the co-array is a "filled" 2D array in the sense that SFPD($\boldsymbol{N}^{(s)}, k_1, k_2$) of the sparse array is completely filled by the dense array sensors for each k_1 and k_2 in the sparse array.

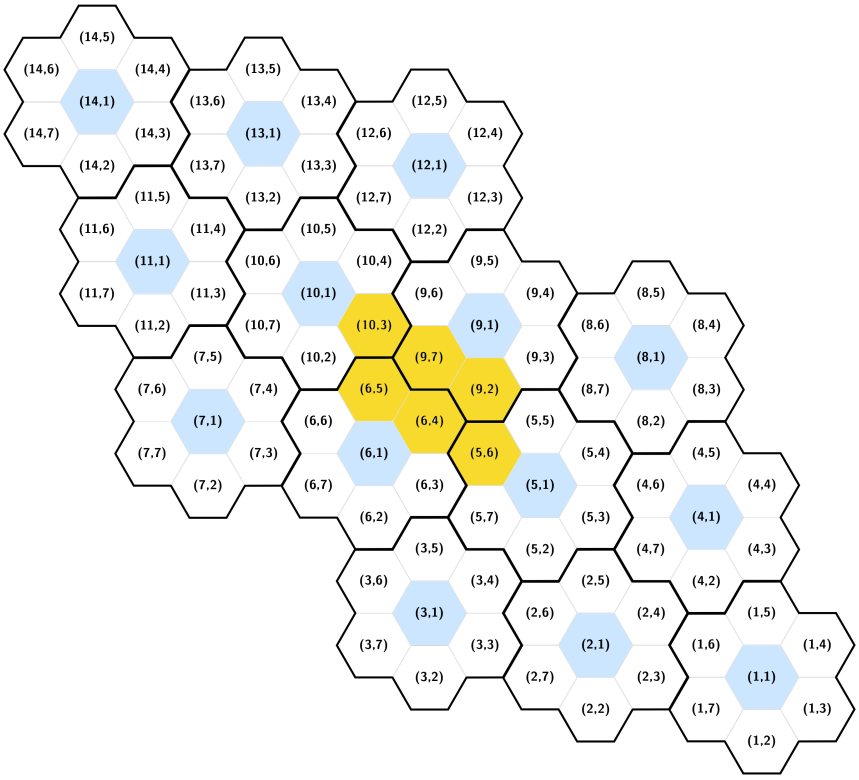

Figure 6.1: Hexagonal cellular system model with nested BSs deployment

From the Definition 5, we also know that the upper left "filled" lattice spans to a parallelogram grid of size $L = MN$, where $M = N_1^{(s)} N_1^{(d)}$ and $N = N_2^{(s)} N_2^{(d)}$.

Every cell has K users. The vector baseband representation of the signals received at the BSs is given as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \tag{6.9}$$

where $\boldsymbol{H}$ is the $L \times LK$ channel transfer matrix

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{a}_0 & 0 & 0 & \dots & 0 & \boldsymbol{b}_0 \\ \boldsymbol{b}_1 & \boldsymbol{a}_1 & 0 & \dots & 0 & 0 \\ 0 & \boldsymbol{b}_2 & \boldsymbol{a}_2 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \boldsymbol{a}_{L-2} & 0 \\ 0 & 0 & 0 & \dots & \boldsymbol{b}_{L-1} & \boldsymbol{a}_{L-1} \end{bmatrix}, \tag{6.10}$$

in which $\boldsymbol{a}_m$ and $\boldsymbol{b}_m$ are $1 \times K$ row vectors denoting the channel coefficients experienced by the K users from the corresponding $\boldsymbol{N}^{(d)}$ and $\boldsymbol{N}^{(s)}$ BSs. It also assumes that the fading coefficients have complex values and are identical and independent distributed (i.i.d.) among different users.

The channel state information (CSI) is assumed available to the joint multicell BSs receivers only, and the channel coefficients are unknown for the transmitter. The transmitters know the channel statistics, which can be viewed as ergodic process. As a result, the users adjust their rates to this ergodic sum-rate capacity. It also assumes that the users cannot cooperate their transmissions, and use Gaussian codebooks so that the transmitted signal $\{[\boldsymbol{x}]_i\}_{i=1}^{LK}$ are i.i.d. zero-mean circularly symmetric Gaussian random variables with variance P. $\boldsymbol{n}$ stands for the zero-mean circularly symmetric additive white Gaussian noise (AWGN) vector, and $E[\boldsymbol{n}\boldsymbol{n}^{\dagger}] = \boldsymbol{I}_L$, where $\boldsymbol{I}_L$ is the $L \times L$ identity matrix. As a result, the power of transmission P is equal to the SNR value in the following derivation.

6.3 Sum-Rate capacity of nested distributed cooperative networks

6.3.1 Invariance in the difference co-array

The concept of invariance is often referred in the algorithms for direction of arrival [90]. Specifically, it means that the array should be divisible into a number of identical subarrays which are the shifted copies of each other.

The maximum number of the subarrays is obtained by shifting the fundamental dense array with successive integer vectors. The improvement of rank is proportional to the number of dense arrays [102], while the degrees of freedom available after spacial smoothing is proportional to the size of the dense array. For constant elements to construct the array system, the best strategy is to make the rank equals to the size of dense array [74]. That is,

$$\boldsymbol{H}_{m,n}(i,j) = \boldsymbol{H}_{\boldsymbol{N}^{(d)}}[m\ n]^T, \tag{6.11}$$

where $-M+1 \leq m \leq M-1$, $-N+1 \leq n \leq N-1$, $0 \leq i < N_1^{(d)}$, $0 \leq j < N_2^{(d)}$.

The received signal by the subarray (m,n) is denoted as $\boldsymbol{y}_{m,n} = \boldsymbol{H}_{m,n}\boldsymbol{x} + \sigma_n^2 \boldsymbol{e}_{m,n}$, where the elements of $\boldsymbol{H}_{m,n}$ are given as

$$[\boldsymbol{H}_{m,n}]_{Nl+i,k} = \boldsymbol{H}_{\boldsymbol{N}^{(d)}} e^{j(\omega_{1k}(l-M+m)+\omega_{2k}(i-N+n))}$$

where $l = 0, \ldots, M$, $i = 0, \ldots, N$, $\omega_{1k} = 2\pi[\cos\theta_k\ \ \sin\theta_k]\boldsymbol{n_1}^{(d)}$, $\omega_{2k} = 2\pi[\cos\theta_k\ \ \sin\theta_k]\boldsymbol{n_2}^{(d)}$, and θ_k is the azimuthal angle of the source k. $\boldsymbol{N}^{(d)} = [\boldsymbol{n_1}^{(d)}\ \boldsymbol{n_2}^{(d)}]$. Then it is ready to derive that

$$\boldsymbol{y}_{m,n} = \boldsymbol{H}_{\boldsymbol{N}^{(d)}} \boldsymbol{\Lambda}_1^m \boldsymbol{\Lambda}_2^n \boldsymbol{x} + \sigma_n^2 \boldsymbol{e}_{m,n} \tag{6.12}$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are $K \times K$ diagonal matrices with (i,i)th element given by $e^{j\omega_1}$ and $e^{j\omega_2}$, respectively. We can now define the autocorrelation of the received signal as

$$\begin{aligned}\boldsymbol{R}_{m,n} \triangleq & \boldsymbol{y}_{m,n}\boldsymbol{y}_{m,n}^{\dagger} \\ = & \boldsymbol{H}_{\boldsymbol{N}^{(d)}}\mathbf{\Lambda}_1^m\mathbf{\Lambda}_2^n\boldsymbol{x}\boldsymbol{x}^{\dagger}(\mathbf{\Lambda}_2^n)^{\dagger}(\mathbf{\Lambda}_1^m)^{\dagger}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger} \\ & + \sigma_n^4\boldsymbol{e}_{m,n}\boldsymbol{e}_{m,n}^{\dagger} + \sigma_n^2\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\mathbf{\Lambda}_1^m\mathbf{\Lambda}_2^n\boldsymbol{x}\boldsymbol{e}_{m,n}^{\dagger} \\ & + \sigma_n^2\boldsymbol{e}_{m,n}\boldsymbol{x}^{\dagger}(\mathbf{\Lambda}_2^n)^{\dagger}(\mathbf{\Lambda}_1^m)^{\dagger}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger} \end{aligned} \tag{6.13}$$

Taking the average of $\boldsymbol{R}_{m,n}$ over all (m,n), we can define the rank-enhanced matrix

$$\hat{R}^2 \triangleq \frac{1}{MN}\sum_{m=0}^{M}\sum_{n=0}^{N}\boldsymbol{R}_{m,n}, \tag{6.14}$$

since it provides the autocorrelations with the freedom of $N^{(s)}N^{(d)}$ BSs by using only $N^{(s)} + N^{(d)}$ BSs. Besides, it can also be shown that

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\mathbf{\Lambda}_1^m\mathbf{\Lambda}_2^n\boldsymbol{x}\boldsymbol{x}^{\dagger}(\mathbf{\Lambda}_2^n)^{\dagger}(\mathbf{\Lambda}_1^m)^{\dagger} = \boldsymbol{R}_{xx}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{R}_{xx} \tag{6.15}$$

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\mathbf{\Lambda}_1^m\mathbf{\Lambda}_2^n\boldsymbol{x}\boldsymbol{e}_{m,n}^{\dagger} = \boldsymbol{R}_{xx}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger} \tag{6.16}$$

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\boldsymbol{e}_{m,n}\boldsymbol{e}_{m,n}^{\dagger} = \boldsymbol{I}_{M\times N} \tag{6.17}$$

Substituting the values from (7.6)-(7.8) in (7.5), we have

Proposition 1. *The covariance matrix of the signal received by a $M \times N$ array of BSs on the lattice $\tilde{\boldsymbol{N}}^{(d)}$ has the same form as $\hat{R}^2$ where*

$$\begin{aligned}\hat{R} &= \frac{1}{\sqrt{MN}}\left(\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{R}_{xx}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger} + \sigma_n^2\boldsymbol{I}_{M\times N}\right) \\ &= \frac{1}{\sqrt{MN}}\left(KP\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}^{\dagger} + \sigma_n^2\boldsymbol{I}_{M\times N}\right). \end{aligned} \tag{6.18}$$

Hence, the cluster of BSs has the degree of freedom as $O(N^{(s)}N^{(d)})$ with $O(N^{(s)}+N^{(d)})$ actual BSs.

6.3.2 Static AWGN channel

Concerning the case of nonfading channels, all fading coefficients are equal to 1. Based on (6.2), it is easy to conclude that (6.1) depends only on the sum of the intra-cell transmit power. Hence, all transmission schemes with equal total intra-cell power achieve the same throughput.

The nonzero entries of $\boldsymbol{H}_L\boldsymbol{H}_L^\dagger$ derived from (6.2) are

$$\left[\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,m} = \sum_{k=1}^{K} |a_{m,k}|^2 |b_{m,k}|^2, \tag{6.19}$$

$$\left[\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,n} = \sum_{k=1}^{K} a_{n,k}^* b_{m,k}, \quad n = (m-1) \bmod L, \tag{6.20}$$

$$\left[\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,n} = \sum_{k=1}^{K} a_{m,k} b_{n,k}^*, \quad n = (m+1) \bmod L. \tag{6.21}$$

As a result, the matrix $\boldsymbol{H}_L\boldsymbol{H}_L^\dagger$ is a circulant matrix with the nonzero row elements $\{K, 2K, K\}$. Based on the eigenvalues of the circulant matrices [88], we can have the following proposition.

Proposition 2. *The uplink average per-cell sum-rate capacity in the absence of fading is*

$$C(P) = \frac{1}{L}\sum_{l=0}^{L} \log_{10}\left(1 + 2KP\left(1 + \cos\left(2\pi\frac{l}{L}\right)\right)\right) \tag{6.22}$$

Substituting (6.22) into (6.4)-(6.6), we could have

Proposition 3. *The spectrum efficiency for uplink channel is characterized by*

$$\frac{E_b}{N_0}_{min} = \frac{\log_e 2}{2}, \ S_{low} = \frac{4}{3}, \ S_{high} = 1. \tag{6.23}$$

6.3.3 Flat-fading channel

The channel fading coefficients are taken as i.i.d. random variables, and their statistics are denoted as

$$m_1 \triangleq E[a_{m,k}] = E[b_{m,k}], \tag{6.24}$$

$$m_2 \triangleq E[|a_{m,k}|^2] = E[|b_{m,k}|^2], \tag{6.25}$$

$$m_4 \triangleq E[|a_{m,k}|^4] = E[|b_{m,k}|^4], \tag{6.26}$$

$$\kappa \triangleq \frac{m_4}{m_2}, \quad \forall m, k, \tag{6.27}$$

to be the mean, second- and fourth-order moments and the kurtosis. The following contents assume that the users have similar large scale path losses to the BSs. Besides, given a rich scattering environment, the small-scale fading processes experienced by the users can be modeled as mutually i.i.d. random processes. For simplicity, it also assumes that all users are received with equal average power from the two BSs based on dense and sparse array, respectively. For the case that all users are simultaneously active, it also assumes that $K \gg 1$. Similarly to the derivation in [79], applying the Strong Law of Large Number (SLLN) to K while keeping the total per-cell transmit power KP constant, the diagonal entries of (6.2) can be represented using (6.24)-(6.27)

$$\lim_{K\to\infty} \left[\frac{1}{K}\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,m} = 2E[|a_{m,k}|^2] = 2m_2 \tag{6.28}$$

$$\lim_{K\to\infty} \left[\frac{1}{K}\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{m,n} = |E[a_{m,k}]|^2 = |m_1|^2 \tag{6.29}$$

$$\lim_{K\to\infty} \left[\frac{1}{K}\boldsymbol{H}_L\boldsymbol{H}_L^\dagger\right]_{n,m} = |E[a_{m,k}]|^2 = |m_1|^2 \tag{6.30}$$

Substituting (6.28)-(6.30) to (6.1), we could have

Proposition 4. *The average per-cell sum-rate capacity while the transmission experiencing Rayleigh fading is*

$$C(P) = \frac{1}{L}\sum_{l=0}^{L-1}\log_{10}\left(1+2KP\left(m_2+|m_1|^2\cos\left(2\pi\frac{l}{L}\right)\right)\right). \tag{6.31}$$

Comparing the capacity (6.31) with (6.22), we can see that the presence of fading enhance the performance in terms of the average per-cell sum-rate capacity. This is because the independence of the two fading processes affecting the signal of each user, as observed by the two receiving BSs. Furthermore, by applying (6.31) to (6.4)-(6.6), we could have

Proposition 5. *For a general fading distribution, $K \gg 1$, and $\forall L \geq 3$, the average per-cell sum-rate capacity is characterized by*

$$\frac{E_b}{N_0}_{min} = \frac{\log_e 2}{2m_2},\ S_{low} = \frac{2}{\frac{\kappa}{2K}+\frac{|m_1|^4}{2m_2}+1},\ S_{high} = 1. \tag{6.32}$$

Specifically for Rayleigh fading, the expressions are

$$\frac{E_b}{N_0}_{min} = \frac{\log_e 2}{2},\ S_{low} = 2,\ S_{high} = 1. \tag{6.33}$$

In the low-SNR region, the minimum transmit E_b/N_0 that enables reliable communications is the same for either intra-cell TDMA, which means that only a single user simultaneously in each cell transmits for a fraction $1/K$ of time with the power P, or in the wideband scenario. It is also identical with or without fading. However, in the presence of Rayleigh fading, employing the wideband scheme with more than two simultaneously active users per cell produces a higher low-SNR slope, and hence a higher spectrum efficiency, as compared to the result for nonfading channels.

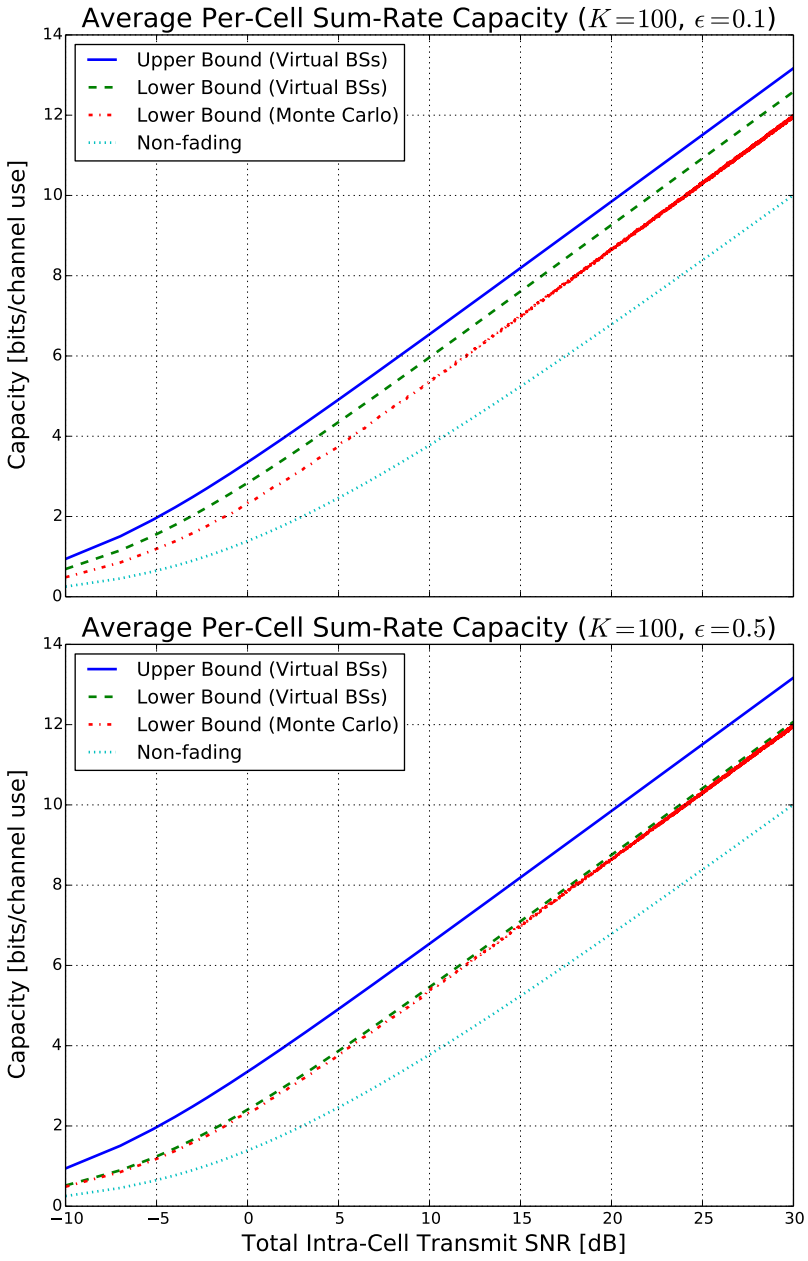

Figure 6.2: Comparison of average per-cell sum-rate capacity for $K = 100$, and $\varepsilon = 0.1$ or $\varepsilon = 0.5$

6.4 Numerical results

Figure 6.2 shows the values of channel capacities in no fading (6.22) and Rayleigh fading (6.31) channels as a function of the total intra-cell transmit power. For com-

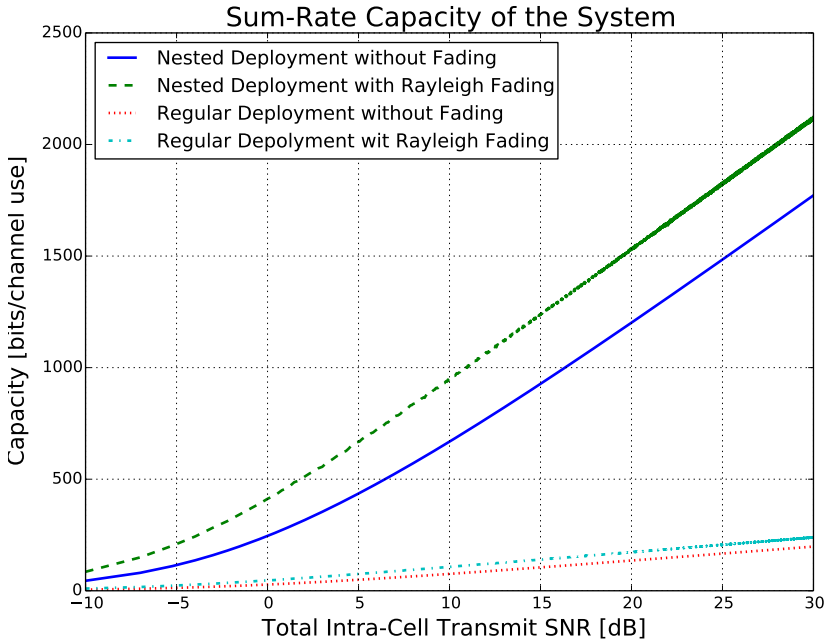

Figure 6.3: Comparison of sum-rate capacity for the system

parison, it also includes the analytical derived upper and lower bounds (6.7) with $K = 100$ users per cell. The upper figure shows the lower bound for $\varepsilon = 0.1$, while the lower figure shows the lower bound for $\varepsilon = 0.5$. Comparing the results of non-fading channels with those of Rayleigh-fading channels, it is obvious that the fading has a positive impact on system performances. Besides, the Monte Carlo simulations have a good match to the analytical lower bound. Furthermore, given a certain number of users in a cell, a higher value of ε implies that more users are likely to experience threshold crossing fading coefficients, so the lower bound, which is derived based on the SLLN, yields a better approximation. This phenomenon is able to be observed by comparing the two figures in Figure 6.2, in which the results of Monte Carlo simulations are closer to the analytical lower bound with $\varepsilon = 0.5$ than it is with $\varepsilon = 0.1$.

Figure 6.3 shows the comparison of sum-rate capacity in the system point of view. The regular deployment stands for the deploying BSs in a common hexagon manner, while the nested deployment is elaborated in the previous contents. Both deployments involve the same number of BSs. Based on the figure, the capacity improvement is substantial. Specifically, the capacity increases by about eight times for non-fading channel, and seven times for the Rayleigh-fading channel.

Figure 6.4 shows the corresponding per-cell spectrum efficiency results, plotted as a function of the system average transmit E_b/N_0. The beneficial effect of fading on system performance is again clearly demonstrated, and a good match to the low-SNR regime characterization of Proposition 5 is observed.

6.5 Concluding discussion

The feasibility and characteristics for the nested distributed cellular network are discussed in this paper based on a traditional hexagonal multi-cell model. The model assumes a modified version of soft-handoff scenario, in which each user simultaneously communicates with two BSs, and at least one of them is the BS from the dense array. Both the non-fading and Rayleigh fading channels are analyzed in terms of the average per-cell sum-rate capacity and spectrum efficiency. It shows that the invariant of difference co-array is valid to analyze the covariance of channel fading coefficients, which plays an important role to derive the cell capacity. The numeric results also support the correctness of the propositions.

A key insight of the present work is that by carefully deploying and scheduling communication between users and different BSs, the cluster of the BSs is able to achieve a much higher capacity compared to the simple collaboration among BSs. On the other hand, there are many open questions in this direction of research, including detailed power allocation and beamforming for the combination of dense array and

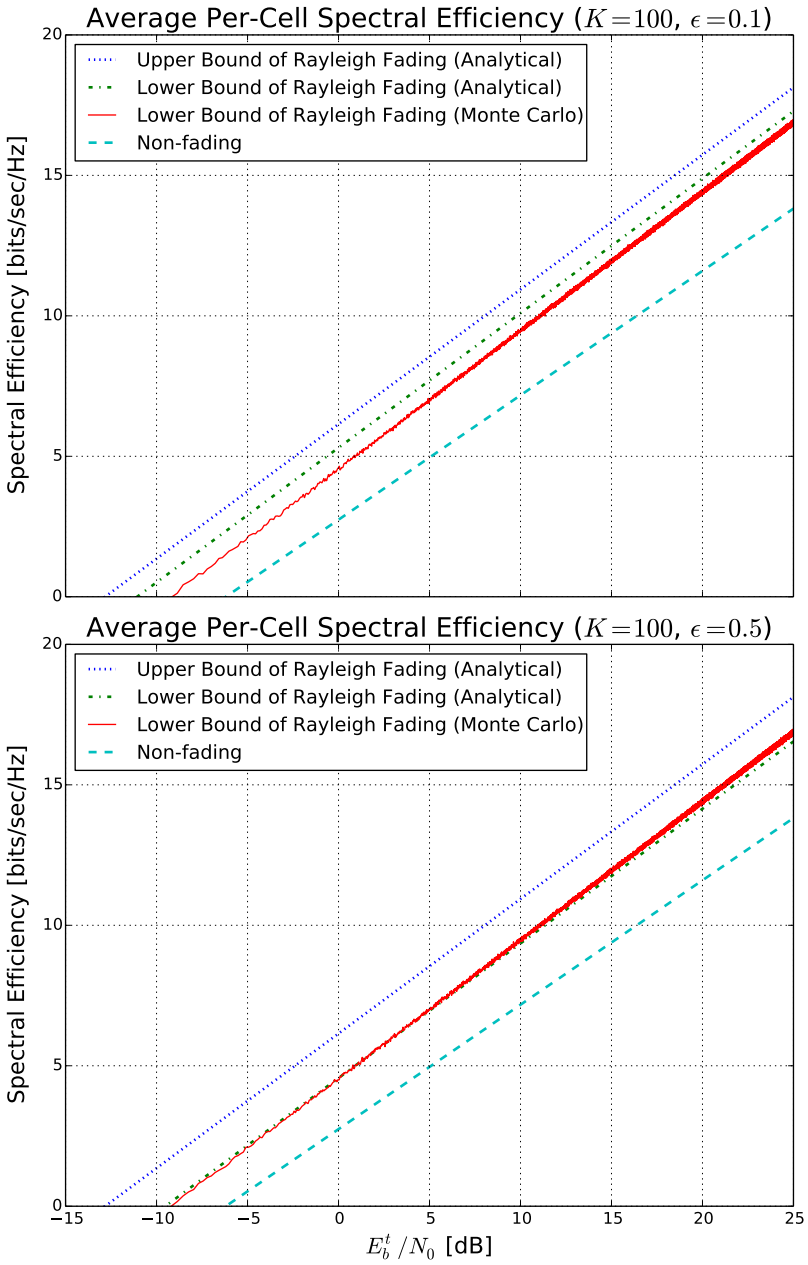

Figure 6.4: Average per-cell spectrum efficiency

sparse array, the interference analysis when this nested distributed cellular network is put into a system at large, and the relation between the uplink and downlink channels.

CHAPTER 7

Ergodic throughput capacity of hybrid wireless networks with nested distributed base stations

7.1 Introduction

The model of hybrid wireless network is proposed to improve the performance of large-scale ad-hoc networks, which has fundamental limitations caused by the interferences and contentions of medium access when the density of nodes per unit becomes large [91]. In contrast, the hybrid networks are able to significantly increase the connectivity by taking advantage of the fixed base stations (BSs) that can be reached in multiple hops [92]. But with the significant growth of data demands by the increasing number of both users and various mobile applications, it requires higher throughput capacity and better utilization of the available spectrum resources. On the other hand, it is appealing to maintain the existing network topology with cooperative transmission schemes to improve the overall capacity of the system.

The concept of distributed antenna systems is originally introduced to enhance the capacity of interference-limited cellular networks by shortening access distance, saving transmit power, and reducing inter-cell interference [93]. It can also be treated as a macroscopic MIMO system [94]. Specifically, the average per-cell capacity of virtual MIMO architecture is analyzed for both uplink [79] and downlink [95]. In the context of the 3GPP LTE-Advanced standard, it is known as coordinated multi-point (CoMP) transmission or reception [96]. Its performance gain is studied in [97] based on the assumption that the information about large-scale fading is known at the transmitter.

Concerning the hybrid wireless network, Gupta and Kumar initiated the study of scaling laws in large ad-hoc wireless networks [98]. The throughput capacity using the successive interference cancellation (SIC) was also investigated in [99] for the fast fading as well as flat fading channels.

One the other hand, however, there is no thorough understanding of the information-theoretic capacity of multi-cell virtual MIMO hybrid network accounting for fading and path loss effects. In this paper, an analytical framework of the hybrid wireless network with distributed base stations is proposed to fill in this blank. Besides, the geographic deployment is also take into consideration. Specifically, the 2D nested array proposed in the paper [73] is used for maximizing the degrees of freedom of the cooperative BSs, so that the system is able to achieve the optimal ergodic throughput capacity given the fixed number of BSs. The assumptions of channels also take reality into consideration: the coherence time is much smaller than the delay spread; both large-scale path loss and small-scale Rayleigh fading are included; and the resulting fast fading channels are independent but not identically distributed.

The rest of this paper is organized as follows. Section 7.2 introduces the theoretical backgrounds about opportunistic communication and 2D nested array, as well as the basic assumptions for the model of nested-distributed BSs. Section 7.3 obtains the invariance of the difference co-array and MIMO capacity to cope with the capacity derivation. In Section 7.4, the ergodic throughput capacity is derived. Finally, we summarize the paper in Section 7.5.

7.2 Preliminaries and model description

In a given cell, it assumes that all inside nodes communicate based on a opportunistic time division multiple access (TDMA) manner. When a node is scheduled for transmission in a particular time slot, a set of nodes from the same cell are selected

to transmit simultaneously with the scheduled source, as long as such transmissions do not impair the achievable transmission rate of the scheduled source. These nodes are referred to as opportunistic sources. All these sources share the entire bandwidth. However, rather than treating the interference from other nodes as noise, the receiver deploys a successive interference cancellation (SIC) technique [100]. That is, after one node is decoded, its signal is stripped away from the aggregate received signal before the next node is decoded.

For any scheduled source s, it assumes that there are K nodes within the same cell, whose Euclidean distance away from destination node d is greater than d_{sd}. That is, we have $d_{0d} \leq d_{1d} \leq \cdots \leq d_{Kd}$, and $d_{kd} \geq (1+\Delta)d_{sd}$, where $k = 0, 1, \ldots, K$, and Δ is a positive constant to model a guard zone [98]. It is worth mentioning the motivation that we choose these farther nodes as opportunistic nodes is to make sure the transmissions from the opportunistic sources do not impair the achievable transmission rate of the scheduled source [91]. The SIC strategy maximizes the sum rate and achieves a set of sum rate satisfying

$$\sum_{k=0}^{K} R_k \leq \log\left(1 + \frac{\sum_{k=0}^{K} \mathcal{P}_k \cdot |h_{kd}|^2}{\sigma^2}\right), \tag{7.1}$$

where R_k denotes the kth node's achievable rate, $\mathcal{P}_k$ is the transmit power from the k_{th} source node, h_{kd} is and the composite channel between the k_{th} source node and the destination node, and σ^2 is the power of AWGN.

7.3 Modeling of hybrid wireless networks

7.3.1 Invariance in the difference co-array

The concept of invariance is often referred in the algorithms for direction of arrival [90]. Specifically, it means that the array should be divisible into a number of identical subarrays which are the shifted copies of each other.

The maximum number of the subarrays is obtained by shifting the fundamental dense array with successive integer vectors. The improvement of rank is proportional to the number of dense arrays [102], while the degrees of freedom available after spacial smoothing is proportional to the size of the dense array. For constant elements to construct the array system, the best strategy is to make the rank equals to the size of dense array [73]. That is,

$$\boldsymbol{H}_{m,n}(i,j) = \boldsymbol{H}_{\boldsymbol{N}^{(d)}}[m\ n]^T, \tag{7.2}$$

where $-M+1 \le m \le M-1$, $-N+1 \le n \le N-1$, $0 \le i < N_1^{(d)}$, $0 \le j < N_2^{(d)}$.

The received signal by the subarray (m,n) is denoted as $\boldsymbol{y}_{m,n} = \boldsymbol{H}_{m,n}\boldsymbol{x} + \sigma_n^2 \boldsymbol{e}_{m,n}$, where the elements of $\boldsymbol{H}_{m,n}$ are given as

$$[\boldsymbol{H}_{m,n}]_{Nl+i,k} = \boldsymbol{H}_{\boldsymbol{N}^{(d)}} e^{j(\omega_{1k}(l-M+m)+\omega_{2k}(i-N+n))},$$

where $l = 0, \ldots, M$, $i = 0, \ldots, N$, $\omega_{1k} = 2\pi[\cos\theta_k\ \ \sin\theta_k]\boldsymbol{n_1}^{(d)}$, $\omega_{2k} = 2\pi[\cos\theta_k\ \ \sin\theta_k]\boldsymbol{n_2}^{(d)}$, and θ_k is the azimuthal angle of the source k. $\boldsymbol{N}^{(d)} = [\boldsymbol{n_1}^{(d)}\ \boldsymbol{n_2}^{(d)}]$. Then it is ready to derive that

$$\boldsymbol{y}_{m,n} = \boldsymbol{H}_{\boldsymbol{N}^{(d)}} \boldsymbol{\Lambda}_1^m \boldsymbol{\Lambda}_2^n \boldsymbol{x} + \sigma_n^2 \boldsymbol{e}_{m,n}, \tag{7.3}$$

where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are $K \times K$ diagonal matrices with (i,i)th element given by $e^{j\omega_1}$ and $e^{j\omega_2}$, respectively. We can now define the autocorrelation of the received signal as

$$\begin{aligned}
\boldsymbol{R}_{m,n} \triangleq & \boldsymbol{y}_{m,n}\boldsymbol{y}_{m,n}^\dagger \\
= & \boldsymbol{H}_{\boldsymbol{N}^{(d)}} \boldsymbol{\Lambda}_1^m \boldsymbol{\Lambda}_2^n \boldsymbol{x}\boldsymbol{x}^\dagger (\boldsymbol{\Lambda}_2^n)^\dagger (\boldsymbol{\Lambda}_1^m)^\dagger \boldsymbol{H}_{\boldsymbol{N}^{(d)}}^\dagger \\
& + \sigma_n^4 \boldsymbol{e}_{m,n}\boldsymbol{e}_{m,n}^\dagger + \sigma_n^2 \boldsymbol{H}_{\boldsymbol{N}^{(d)}} \boldsymbol{\Lambda}_1^m \boldsymbol{\Lambda}_2^n \boldsymbol{x}\boldsymbol{e}_{m,n}^\dagger \\
& + \sigma_n^2 \boldsymbol{e}_{m,n}\boldsymbol{x}^\dagger (\boldsymbol{\Lambda}_2^n)^\dagger (\boldsymbol{\Lambda}_1^m)^\dagger \boldsymbol{H}_{\boldsymbol{N}^{(d)}}^\dagger .
\end{aligned} \tag{7.4}$$

Taking the average of $\boldsymbol{R}_{m,n}$ over all (m,n), we can define the rank-enhanced matrix

$$\hat{R}^2 \triangleq \frac{1}{MN} \sum_{m=0}^{M} \sum_{n=0}^{N} \boldsymbol{R}_{m,n}, \tag{7.5}$$

Besides, it can also be shown that

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\boldsymbol{\Lambda}_1^m\boldsymbol{\Lambda}_2^n\boldsymbol{x}\boldsymbol{x}^\dagger(\boldsymbol{\Lambda}_2^n)^\dagger(\boldsymbol{\Lambda}_1^m)^\dagger = \boldsymbol{R}_{xx}\boldsymbol{H}^\dagger_{\boldsymbol{N}^{(d)}}\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{R}_{xx} \tag{7.6}$$

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\boldsymbol{\Lambda}_1^m\boldsymbol{\Lambda}_2^n\boldsymbol{x}\boldsymbol{e}^\dagger_{m,n} = \boldsymbol{R}_{xx}\boldsymbol{H}^\dagger_{\boldsymbol{N}^{(d)}} \tag{7.7}$$

$$\sum_{m=0}^{M}\sum_{n=0}^{N}\boldsymbol{e}_{m,n}\boldsymbol{e}^\dagger_{m,n} = \boldsymbol{I}_{M\times N} \tag{7.8}$$

Substituting the values from (7.6)-(7.8) in (7.5), we have

Proposition 6. *The covariance matrix of the signal received by a $M \times N$ array of BSs on the lattice $\tilde{\boldsymbol{N}}^{(d)}$ has the same form as $\hat{R}^2$ where*

$$\begin{aligned}\hat{R} &= \frac{1}{\sqrt{MN}}\left(\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{R}_{xx}\boldsymbol{H}^\dagger_{\boldsymbol{N}^{(d)}} + \sigma_n^2\boldsymbol{I}_{M\times N}\right)\\ &= \frac{1}{\sqrt{MN}}\left(KP\boldsymbol{H}_{\boldsymbol{N}^{(d)}}\boldsymbol{H}^\dagger_{\boldsymbol{N}^{(d)}} + \sigma_n^2\boldsymbol{I}_{M\times N}\right).\end{aligned} \tag{7.9}$$

Hence, the cluster of BSs has the degree of freedom as $N^{(s)}N^{(d)}$ with $N^{(s)} + N^{(d)}$ physical BSs.

As a result, the nested distributed base stations (NDBS) is able to acquire densely deployed $O(N^{(s)}N^{(d)})$ virtual BSs via the difference co-array with $O(N^{(s)} + N^{(d)})$ actual BSs. In the following sections, these BSs simulate the multiple antennas in the communication providing diversity gain and increasing reliability. For a target cell, it can be served by the BSs not only from sparse array, but from the dense array sharing the same frequency band.

7.3.2 Formulate NDBS as MIMO system

We denote N_{BS} as the antennas equipped in each BS, and assume that each user node has single antenna ($N_{nd} = 1$). A total bandwidth of W Hz for data transmission is further split into W_u Hz for uplink, W_d Hz for downlink, and $W = W_u + W_d$. There are three phases for the communication. During the uplink, the source node transmits

data to all the NDBS under the existence of opportunistic nodes in the same cell. Then, the collaborative NDBS decode the received data and send them to the NDBS which are serving the destination cell via the wired network. Finally, the data is transmitted to the destination node by cooperative NDBS via downlink.

For uplink, the received signal at the NDBS is described as:

$$\mathbf{y_u} = \sum_{k=0}^{\kappa} \sqrt{\mathcal{P}_u}\mathbf{H}_k^H x_k + \mathbf{n}, \tag{7.10}$$

where $\mathcal{P}_u$ is the average transmit power for source nodes and opportunistic nodes; κ is the number of opportunistic nodes which simultaneously generate traffic to m NDBS; $\mathbf{H}_k$ is the composite channel matrix from the k_{th} node to m NDBS; x_k represents the normalized transmit symbol sent by the k_{th} node; the noise $\mathbf{n}$ is of AWGN $\mathcal{CN}(0, \sigma_n^2 I_{N \cdot N_{BS}})$, and the inter-cell interference which can be modeled as color Gaussian noise [99].

Particularly, $\mathbf{H}_k$ consists N independent subchannel matrices to each NDBS:

$$\mathbf{H}_k = [\mathbf{H}_1^k \quad \mathbf{H}_2^k \ldots \mathbf{H}_n^k \ldots \mathbf{H}_m^k], \tag{7.11}$$

where $\mathbf{H}_n^k : \mathcal{C}^{1 \times N_{BS}}$ represents the composite fast fading channel matrix from the k_{th} node to the n_{th} NDBS:

$$\mathbf{H}_n^k = h_n^{ls}\mathbf{H}_{n,k}^{ss}, \quad n = 1, 2, \ldots m, \tag{7.12}$$

where $h_n^{ls} = \frac{e^{-\gamma d_n/2}}{d_n^{\alpha/2}}$ denotes the large-scale path loss to the n_{th} NDBS, in which the distance between the n_{th} NDBS and the source node is d_n, and γ and α are the absorption constant of the attenuation and path-loss exponent. We assume all the nodes are uniformly distributed in the two-dimensional plane. $\mathbf{H}_{n,k}^{ss} : \mathcal{C}^{1 \times N_{BS}}$ is the small-scale Rayleigh fading channel matrix, in which all the entries are standard normal distributed.

Similarly, the received signal for the downlink is expressed as:

$$y_d = \sqrt{\frac{\mathcal{P}_d}{N_{BS}}}\mathbf{H}^H\mathbf{x} + n'. \tag{7.13}$$

$\mathcal{P}_d$ is the transmit power at each NDBS. $\mathbf{x} : \mathcal{C}^{mN_{BS}\times 1}$ represents the normalized transmit symbol matrix from each transmit antenna. $\mathbf{H}$ is the composite channel matrix from m NDBS to the destination node, which can be further separated as

$$\mathbf{H} = [\mathbf{H}_1 \quad \mathbf{H}_2 \ldots \mathbf{H}_n \ldots \mathbf{H}_m]^T. \tag{7.14}$$

$\mathbf{H}_n : \mathcal{C}^{N_{BS}\times 1}$ represents the composite fast fading channel matrix from the n_{th} NDBS to the destination code:

$$\mathbf{H}_n = h_n^{ls}\mathbf{H}_n^{ss}, \quad n = 1, 2, \ldots m, \tag{7.15}$$

where $\mathbf{H}_n^{ss} : \mathcal{C}^{N_{BS}\times 1}$ is the small-scale Rayleigh fading channel matrix, in which all the entries have standard normal distribution.

7.4 Ergodic throughput capacity

7.4.1 The number of nodes per cell

The following lemma shows that there are $\Theta(\frac{n}{b})$ nodes within each cell.

Proposition 7. *Consider a hexagonal network model consists of n nodes and b base stations, both of which are uniformly distributed. For $b = O(\frac{n}{\log n})$ and $n \to \infty$, the number of nodes in each cell is bounded by $\Theta(\frac{n}{b})$.*

Proof. Let event A denote a Bernoulli event that a particular node i, $1 \leq i \leq n$, will fall into a particular cell of area c^2. Because nodes are placed uniformly on the network, it is clear that probability of event A is $P_A = \frac{n/b}{n} = \frac{1}{b}$. Therefore, the

number of nodes, n_c, has a binomial distribution with parameters (P_A, n). Using Chernoff bound, we have

$$Pr(n_c > k_1\frac{n}{b}) \leq \frac{E\{\exp(n_c)\}}{\exp(\frac{k_1 n}{b})}$$

where k_1 is a constant. Since $E\{\exp(n_c)\} = (1+(e-1)P_A)^n \leq \exp[(e-1)\frac{n}{b}]$ because $1+x \leq exp(x)$, we get

$$Pr(n_c > k_1\frac{n}{b}) \leq \ \exp\left\{ -\frac{n}{b}[k_1 - (e-1)]\right\}. \tag{7.16}$$

As long as $k_1 > e-1$, we know by the union bound that Pr(some cells have more than $\frac{k_1 n}{b}$ nodes) converges to zero as n tends to infinity. Similarly,

$$Pr(n_c < k_2\frac{n}{b}) \leq \frac{E\{\exp(-n_c)\}}{\exp(\frac{-k_2 n}{b})}$$

where k_2 is also a constant. Since $E\{\exp(-n_c)\} = (1+(e^{-1}-1)P_A)^n \leq \exp[(e^{-1}-1)\frac{n}{b}]$, we obtain

$$Pr(n_c < k_2\frac{n}{b}) \leq \ \exp\left\{ -\frac{n}{b}[(1-e^{-1}) - k_2]\right\}. \tag{7.17}$$

As long as $k_2 < 1-e^{-1}$, we know by the union bound that Pr(some cells have less than $\frac{k_2 n}{b}$ nodes) converges to zero as n tends to infinity. Hence, it is concluded that each cell contains $\Theta(\frac{n}{b})$ nodes and we complete the proof. □

7.4.2 Uplink ergodic throughput capacity

The ergodic capacity for the fasting fading channel is defined as the ensemble average of channel capacity over all possible channel realizations [101]. If we assume the channel state information is known only at the receiver, combining with the received signal model of the uplink phase in (7.10), the ergodic capacity with opportunistic communications and SCI scheme is expressed as:

$$R^u = \mathbb{E}\left[\log\left(1+\frac{\sum_{k=0}^{\kappa}\mathcal{P}_u\|\mathbf{H}_k\|^2}{I_u+\sigma_n^2}\right)\right]. \tag{7.18}$$

Given the properties of the composite uplink channel matrix in (7.11), the ergodic capacity of the source node and opportunistic nodes is further derived as:

$$R^u = \mathbb{E}\left[\log\left(1+\gamma_u \sum_{n=1}^{m}\beta_n \sum_{k=0}^{\kappa}\|\mathbf{H}_{n,k}^{ss}\|^2\right)\right], \tag{7.19}$$

where $\gamma_u = \frac{\mathcal{P}_u}{I_u+\sigma_n^2}$ is the ratio of signal to noise plus interference (SINR), and $\beta_n = (h_n^{ls})^2$.

Since the entries of $\mathbf{H}_{n,k}^{ss}$ are independent identically distributed (*i.i.d.*) complex Gaussian random variables, $\|\mathbf{H}_{n,k}^{ss}\|^2$ follows Chi-square distribution with $2N_{BS}$ degrees of freedom. Furthermore, we denote

$$\varphi = \sum_{n=1}^{m}\varphi_n = \sum_{n=1}^{m}\beta_n \sum_{k=0}^{\kappa}\|\mathbf{H}_{n,k}^{ss}\|^2. \tag{7.20}$$

We can get $\varphi_n \sim \Gamma(\kappa N_{BS}, 2\beta_n)$. As a result, φ can be approximated as another Gamma distribution as $\Gamma(k_\varphi, \theta_\varphi)$ using the second order moment matching proposed in [95], with the same first and second moments as:

$$k_\varphi = \frac{\kappa N_{BS}(\sum_{n=1}^{m}\beta_n)^2}{\sum_{n=1}^{m}(\beta_n)^2} \quad \text{and} \quad \theta_\varphi = \frac{2\sum_{n=1}^{m}(\beta_n)^2}{\sum_{n=1}^{m}\beta_n}, \tag{7.21}$$

with the mean and variance

$$\mu_\varphi = 2\kappa N_{BS}\sum_{n=1}^{m}\beta_n \quad \text{and} \quad \sigma_\varphi^2 = 4\kappa N_{BS}\sum_{n=1}^{m}\beta_n^2 \tag{7.22}$$

For high SINR, we have

$$\begin{aligned} R^u &\approx \mathbb{E}\left[\log\left(\gamma_u \sum_{n=1}^{m}\beta_n \sum_{k=0}^{\kappa}\|\mathbf{H}_{n,k}^{ss}\|^2\right)\right] \\ &= \log_2\gamma_u + \log_2 e\cdot\psi(k_\varphi) + \log_2\theta_\varphi \\ &\overset{(a)}{=} \log_2(\gamma_u\theta_\varphi) + \log_2 e\left[\ln k_\varphi + \frac{1}{k_\varphi} + O\left(\frac{1}{k_\varphi^2}\right)\right] \end{aligned} \tag{7.23}$$

where $\psi(.)$ in is the digamma function, and its asymptotic approximation $\psi(x) = \ln x + \frac{1}{x} + O\left(\frac{1}{x}\right)$ is used in (a).

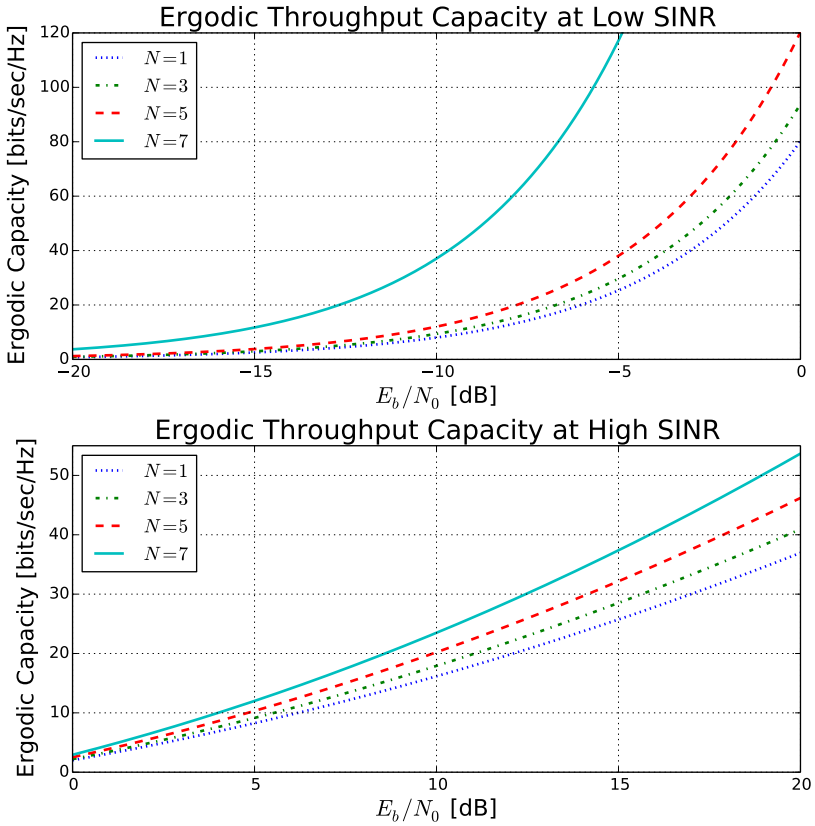

Figure 7.1: Ergodic Capacity in the uplink phase with $N_{BS} = 2$

In the case of low SINR,

$$R^u \approx \mathbb{E}\left[\log_2 e \cdot \gamma_u \sum_{n=1}^{m} \beta_n \sum_{k=0}^{\kappa} \left\|\mathbf{H}_{n,k}^{ss}\right\|^2\right]$$
$$= \log_2 e \cdot \gamma_u \cdot 2\kappa N_{BS} \sum_{n=1}^{m} \beta_n. \tag{7.24}$$

Because $\kappa = \Theta(\frac{mn}{b})$, the uplink ergodic capacity is upper bounded by

$$R^u = O\left(\log\left(\frac{mnN_{BS}}{b}\right)\right). \tag{7.25}$$

Theorem 7. *For a hybrid wireless network of n nodes and m distributed base stations with N_{BS} transmit antennas equipped at each base station, if the total number of base*

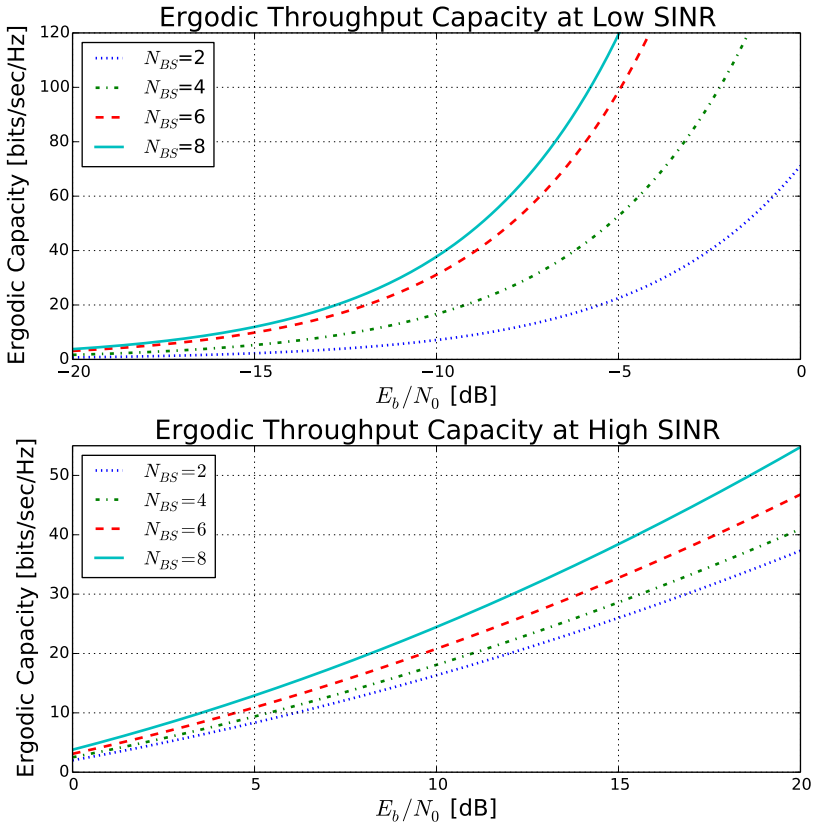

Figure 7.2: Ergodic Capacity in the uplink phase with $m = 2$

stations $b = o(\frac{n}{\log n})$*, the uplink ergodic throughput capacities over Rayleigh fading channels are*

$$T^u_{high}(m,n,b) = O\left(\log\left(\frac{mnN_{BS}}{b}\right)W_u\right) \quad bit/s, \tag{7.26}$$

$$T^u_{low}(m,n,b) = O\left(\frac{mnN_{BS}}{b}W_u\right) \quad bit/s, \tag{7.27}$$

for high SINR and low SINR respectively.

Numerical simulations are presented to validate this theorem. Figure 7.1 shows the ratio between ergodic capacity and the capacity under AWGN at high SNR with $m_{BS} = 4$ antennas at NDBS ($\kappa = 50$), in which $m = 1$ indicates no cooperation between base stations. By comparing the performance, one can attribute the capacity

improvement to the number of NDBS involved in (m). Similarly, Figure 7.2 illustrates uplink performance at high SINR with $m = 9$ NDBS. It is obvious that the ergodic capacity scales in $O\big(N_{BS}\big)$.

7.4.3 Downlink ergodic throughput capacity

Given the received signal model as in (7.13), the ergodic capacity of the downlink phase under infrastructure mode is:

$$\begin{aligned} R^d &= \mathbb{E}\left[\log\left(1+\frac{\mathcal{P}_d\|\mathbf{H}\|^2}{N_{BS}(I_d+\sigma_n^2)}\right)\right] \\ &= \mathbb{E}\left[\log\left(1+\frac{\gamma_d}{N_{BS}}\cdot\sum_{n=1}^{m}(h_n^{ls})^2\|\mathbf{H}_n^{ss}\|^2\right)\right]. \end{aligned} \tag{7.28}$$

Furthermore, we denote

$$\chi = \sum_{n=1}^{m}\chi_n = \sum_{n=1}^{m}(h_n^{ls})^2\|\mathbf{H}_n^{ss}\|^2. \tag{7.29}$$

Hence, $\chi_n \sim \Gamma(N_{BS}, 2\beta_n)$, and χ can be approximated by Gamma distribution $\Gamma(k_\chi, \theta_\chi)$ with

$$k_\chi = \frac{N_{BS}(\sum_{n=1}^{m}\beta_n)^2}{\sum_{n=1}^{m}(\beta_n)^2} \quad \text{and} \quad \theta_\chi = \frac{2\sum_{n=1}^{m}(\beta_n)^2}{\sum_{n=1}^{m}\beta_n} \tag{7.30}$$

as the first and second order moments,

$$\mu_\chi = 2N_{BS}\sum_{n=1}^{m}\beta_n \quad \text{and} \quad \sigma_\chi^2 = 4N_{BS}\sum_{n=1}^{m}\beta_n^2, \tag{7.31}$$

as the mean and variance, where $\beta_n = e^{-\gamma d_n} d_n^{-\alpha}$.

As a result, the ergodic capacity is derived from (7.28) as

$$R^d \approx \log_2\frac{\gamma_d\theta_\chi}{N_{BS}} + \left[\ln k_\chi + \frac{1}{k_\chi} + O\left(\frac{1}{k_\chi^2}\right)\right]\log_2 e \tag{7.32}$$

for high SINR, and for low SINR, it is

$$R^d = 2\gamma_d\left(\sum_{n=1}^{m}\beta_n\right)\log_2 e \tag{7.33}$$

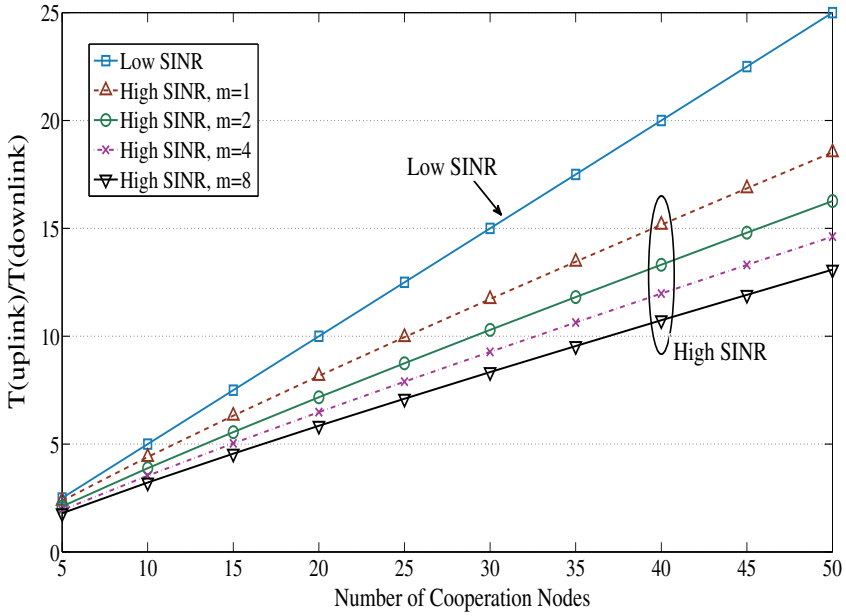

Figure 7.3: Ratio between uplink and downlink ergodic throughput capacity

Hence, at high SINR, the downlink transmission rate is $\Theta(\log(\frac{mn}{b}))$, and the rate is $\Theta(\frac{mn}{b})$ at low SINR. Because the badwidth limitation of the downlink is $\Theta(\frac{b}{n}W_2)$, the ergodic throughput capacity is summarized as follow.

Theorem 8. *For a hybrid wireless network of n nodes and N distributed base stations over Rayleigh fading channels, if the total number of base stations $b = o(\frac{n}{\log n})$, the per-node downlink ergodic throughput capacity at high SINR is*

$$T^d_{high}(m,n,b) = O\big(\frac{mb}{n}\log(\frac{mn}{b})W_d\big) \quad bit/s, \tag{7.34}$$

and at low SINR,

$$T^d_{low}(m,n,b) = O\big(mW_d\big) \quad bit/s. \tag{7.35}$$

Figure 7.3 further illustrates the asymptotic behaviors of the derived ergodic capacity. Its vertical axis is the ratio between the uplink and the downlink capacity. The advantage of introducing opportunistic sources is very obvious. At both low SINR (5dB) and high SINR (15 dB), the uplink throughputs exceed the downlink

counterparts as the opportunistic nodes κ increase. However, with the m increasing, the differences between the uplink and downlink capacity decreases.

7.5 Conclusion

In this paper, we demonstrate the advantages of NDBS in hybrid wireless networks based on the information-theoretic analysis. To maximize the degrees of freedom offered by the existing base stations, we introduce two-dimensional nested array, and derive the invariance in the difference co-array so that the virtual BSs are able to be much denser than their physical deployment. Based on this premise, we interpret the hybrid wireless network as a MIMO system with cooperative BSs. The analysis of the ergodic throughput capacity shows that the improvement of performances is linearly related to the number of collaborative BSs as well as the number of the antennas in each BS. It is concluded that, with opportunistic sources, although the capacity in uplink is more robust to fading as wireless nodes increases, its counterpart in downlink is also able to be improved by inviting more BSs as transmitters.

CHAPTER 8

Conclusion and future works

8.1 Conclusion

In this dissertation, the theories of sparse sampling and array are derived for both signal processing and information theory in wireless communication. Based on the research results, the basis of sparse sampling is not necessarily confined to the WWS signal. Combined with the simulation for non-stationary signal, the areas of implementing sparse sampling is greatly broadened. Besides, the utilization of co-prime and nested pairs is also expanded beyond calculating autocorrelation. HOS has been proved to be useful in decades ago, but the prohibited computation complexity stops it from widely used. The research in this dissertation introduces a novel approach to decrease the complexity while still maintains all of the characteristic merits.

Furthermore, this dissertation also develops the information theoretical analysis for cooperative wireless communication system given nested distributed base stations. A key insight of the present work is that by carefully deploying and scheduling communication between users and different BSs, the cluster of the BSs is able to achieve a much higher capacity compared to the simple collaboration among BSs. On the other hand, there are many open questions in this direction of research, including detailed power allocation and beamforming for the combination of dense array and sparse array, the interference analysis when this nested distributed cellular network is put into a system at large, and the relation between the uplink and downlink channels.

8.2 Future works

8.2.1 Channel capacity under sub-Nyquist coprime sampling

The study of capacity of analog Gaussian channels and capacity-achieving transmission strategies was pioneered by Shannon [50]. These results have provided fundamental insights for modern communication system design. Shannons work focused on capacity of analog channels sampled at or above twice the channel bandwidth. However, in practice, the Nyquist rate may be excessive for perfect reconstruction of signals that possess certain structures known a priori. On the other hand, the hardware and power limitations may preclude sampling at the Nyquist rate for a wideband communication system. For example, the newly developed LTE-Advanced standard allows carrier aggregation, where multiple subbands over a wide spectrum are aggregated and jointly used for transmission to/from a single terminal. Then, the sampling rate requirement for perfect recovery is the sum of the subband bandwidths termed the lamdau rate [51].

This motivates the exploration of the effects of sub-Nyquist sampling upon the capacity of an analog Gaussian channel, and the fundamental capacity limits that result when considering general sampling methods that include irregular nonuniform sampling.

When the channel or signal structure is unknown, the blind sub-Nyquist sampling approaches have be proposed to exploit the structure of various classes of input signals based on sampling with modulation and filter banks [52]. The key step is to scramble spectral contents from different subbands through the modulation operation. This procedure is also a characteristic of a general class of realizable nonuniform sampling techniques applied in practice.

One of the important characteristics for the subsampled channel is the non-invertibility of the prefiltering operation, i.e., we cannot recover the analog channel output from sub-Nyquist samples. The aliased SNR is a convex combination of SNRs at all aliased branches, indicating that the impulse responses of the prefilter play the role of "weighting" different branches. As in maximum-ratio combining (MRC), the frequencies with larger SNRs should be given larger weight, while those that suffer from poor channel conditions should be suppressed. Then, the problem of finding optimal prefilters corresponds to joint optimization over all input and filter responses.

At the transmitter side, however, although MRC maximizes the combiner SNR for a MISO channel, it is suboptimal for the joint optimization problem compared with selection combining [53], which sets $S(f - lf_s)$ to one for some $l = l_0$ and all of the other frequencies to be zero. It precludes the undesired effects of noise from low SNR frequencies, which is crucial in maximizing data rate. Besides, given the transmitted signal is controlled, the selective combining prefiltering is able to generate alias-free channel as well as suppress out-of-band noise.

The capacity with sampling rate under filter- and modulation-bank sampling are not monotonously increasing [54]. This indicates that more sophisticated sampling techniques, adaptive to the channel response and the sampling rate, are necessary to maximize the capacity under sub-Nyquist rate constraints, including both uniform and nonuniform sampling. Then, the following questions are which sampling method can best exploit the channel structure and maximize sampled capacity under a given sampling rate constraint. Besides, another question is that whether exists a capacity upper bound over a general class of sub-Nyquist sampling systems.

I propose to characterize sampled channel capacity as a function of different coprime sampling rate, thereby forming a new connection between sampling theory and information theory. Besides, the study could indicate how the capacity of a

sampled analog channel is affected by reduced sampling rate and identify optimal sampling structures with certain choices of coprime pair given channel condition.

8.2.2 Adaptive multi-level nested sampling

Channel estimation is a well-studied problem in the fileds of telecommunication and signal processing. It has become a popular topic again due to its important implementation in the modern wireless communication system. Combination of Space-Time Block Code (STBC) with Orthogonal Frequency Division Multiplexing (OFDM) has the potential to approach the information theoretical capacity limit of Multiple Input Multiple Output (MIMO) channels [55]. However, at the receiver side, most space-time equalizer require the knowledge of the Channel State Information (CSI) to recover the transmitted data. This information is usually obtained through training the coherent Maximum-Likelihood (ML) receiver [56]. The drawbacks of training-based approaches and differential schemes have motivated an increasing interest in the development of blind channel estimation algorithms for STBC systems.

Despite the high performances of ML algorithm, their computational costs become prohibitive for high-order modulations. In the case of BPSK or QPSK constellations, the blind-ML detection can be simplified to a Boolean Quadratic Program (BQP) [57]. For more general settings, iterative procedure can be employed to avoid the computational complexity of the ML approach. These include the Cyclic ML [56] and the Expectation-Maximization (EM) [58] algorithms. However, these iterative methods require a careful initialization of the channel and/or symbols. In particular, a poor initialization can strongly affect the Symbol-Error Rate (SER) performance. To avoid these drawbacks, several authors have investigated the use of sub-space [59] or second-order statistics [60, 61] approaches. However, excluding some specific low-rate codes, these approaches fail to extract the channel in a full-blind context [59–61].

Several approach have been proposed in literature to solve this problem, including the transmission of a short training sequence [59] or the use of precoders [61]. However, these semi-blind methods cannot be employed in a non-cooperative scenario since they require modification of the transmitter.

On the other hand, although HOS is able to avoid these limitations, based on the previous [62], the original HOS algorithm is computational expensive while the PCS-based HOS has a very slow speed of convergence.

REFERENCES

[1] S. W. Lang, G. L. Duckworth, and J. H. McClellan, "Array design for MEM and MLM array processing," in *Proceedings IEEE IEEE International Conference on Acoustics, Speech and Signal Processing*, vol.6, pp. 145148, April 1981.

[2] T. Hastie and R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 3rd ed., Springer, Feb. 2009.

[3] A. Moffet, "Minimum Redundancy Linear Arrays," *IEEE Transctions on Antennas Propagation*, vol. 16, pp. 172175, Mar. 1968.

[4] S. D. Bedrosian, "Nonuniform Linear Arrays: Graph-theoretic Approach to Minimum Redundancy," *Proceedings of IEEE*, vol. 74, pp. 1040-1043, July 1986.

[5] D. Pearson, S. U. Pillai, and Y. Lee, "An Algorithm for Near-optimal Placement of Sensor Elements," *IEEE Transctions on Information Theory*, vol. 36, pp. 12801284, November 1990.

[6] P. Pal and P. P. Vaidyanathan, "Nested Arrays: A Novel Approach to Array Processing with Enhanced Degrees of Freedom," *IEEE Transctions on Signal Processing*, vol. 58, no. 8, pp. 41674181, August 2010.

[7] P. Pal and P. P. Vaidyanathan, "Coprime Sampling and the MUSIC Algorithm," *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop*, pp. 289-294, 4-7 January 2011.

[8] J. Chen, Q. Liang, J. Wang, H. A. Choi, "Spectrum Efficiency of Nested Sparse Sampling," *Proc. of 7th Int. Conf. Wireless Algorithms, Systems, and Applications*, vol. 7405, pp. 574-583, Yellow Mountains, China, August, 2012.

[9] G. Li, J. Xu, Y.-N. Peng, and X.-G. Xia, "Location and Imaging of Moving Targets Using Nonuniform Linear Antenna Array SAR," *IEEE Transctions on Aerospace Electronic Systems*, vol. 43, no. 3, pp. 12141219, July 2007.

[10] Z. A. Lomnicki and S. K. Zaremba, "On the Estimation of Autocorrelation in Time Series," *The Annals of Mathematical Statistics*, Vol. 28, No. 1, pp. 140-158, March 1957.

[11] M. S. Bartlett, "On the Theoretical Specification and Sampling Properties of Autocorrelation Time Series," *Supplement to the Journal of the Royal Statistical Society*, vol. 8, no. 1, pp. 27-41, 1946.

[12] F. H. C. Marriott and J. A. Pope, "Bias in the Estimation of Autocorrelations," *Biometrika*, Vol. 41, No. 3/4, pp. 390-402, December 1954.

[13] S. M. Kay, "The Effect of Sampling Raate on Autocorrelation Estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 20, no. 4, pp. 859-867, August 1981.

[14] M. Skolnik, *Introduction to Radar Systems*, McGraw-Hill Science, Engineering, and Math, 3rd ed, pp. 276-340, December 2002.

[15] M. A. Richards, *Fundamentals of Radar Signal Processing*, McGraw-Hill, 1st ed, pp. 159-233, June 2005.

[16] T. Nagell, *Introduction to Number Theory*, American Mathematical Society, 2nd Reprint edition, June 2001.

[17] G. B. Giannakis and M. K. Tsatsanis, "Signal detection and classification using matched filtering and higher order statistics," *IEEE Trans. Acous. Speech and Signal Processing*, vol. 38, no. 7, pp. 1284-1296, July 1990.

[18] S. Barbarossa and G. B. Giannakis, "Product high-order ambiguity function for multicomponent polynomial-phase signal modeling," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 691-708, March 1998.

[19] K. S. Lii, "Nonlinear systems and higher-order statistics," *IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 1-8, June 1993.

[20] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 3, pp. 360-377, March 1989.

[21] J. M. Mendel, "Use of higher-order statistics in signal processing and system theory: An update," in *Proc. of SPIE Conf. on Adv. Algo. and Arch. for Signal Processing*, San Diego, CA, pp. 126-144, Feb. 1988.

[22] M. K. Tsatsanis and G. B. Giannakis, "Object detection and classification using matched filtering and higher ordcr statistics," in *Proc. of 6th Multi. Signal Processing Workshop*, pp. 32-33, Sep. 1989.

[23] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," in *Proc. of IEEE*, val. 79, no. 3, pp. 278-305, March 1991.

[24] K. S. Lii and K. N. Helland, "Cross-bispectrum computation and variance estimation," *ACM Trans. Mathematical Software*, vol. 7, no. 3, pp. 284-294, Sep. 1981.

[25] Q. Wu and Q. Liang, "Co-prime sampling for nonstationary signal in radar signal processing," *EURASIP Journal on Wireless Communications and Networking*, doi:10.1186/1687-1499-2013-58, March 2013.

[26] X. Xia, "On estimation of multiple frequencies in undersampled complex valued waveforms," *IEEE Trans. Signal Processing*, vol. 47, no. 12, pp. 3417-3419, Dec. 1999.

[27] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Proc.*, vol. 59, pp. 573-586, Feb. 2011.

[28] —-, "System identification with sparse co-prime sensing," *IEEE Sig. Proc. Letters*, vol. 17, no. 10, pp. 823-826, Oct. 2010.

[29] —-, "Theory of sparse co-prime sensing in multiple dimensions," *IEEE Trans. Signal Processing*, vol. 59, no. 8, pp. 3592-3608, Aug. 2011.

[30] G. Everest, *An Introduction to Number Theory*, Springer, Nov. 2010.

[31] S. Winograd, "On computing the discrete Fourier transform," *Mathematics of Computation*, vol. 32, no. 1, pp. 175-199, Jan. 1978.

[32] J. J. Kormylo and J. M. Mendel, "Identifiability of nonminimum phase linear stochastic systems," *IEEE Trans. on Auto. Control*, vol. 28, no. 12, pp. 1081-1090, Dec. 1983.

[33] C. S. Burrus and P. W. Eschenbacher, "An in-place, in-order prime factor FFT algorithm," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 29, no. 4, pp. 806-817, Aug. 1981.

[34] D. S. Baum and J. Hansen, "An interim channel model for beyond-3G systems," *IEEE Vehicular Tech. Conf.*, vol. 5, pp. 3132-3136, May 2005.

[35] J. K. Tugnait, "Approaches to FIR system identification with noisy dsing higher-Order statistics," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1307-1317, July 1990.

[36] S. S. Ghassemzadeh, R. Jana, C. W. Rice, and W. Turin, "Measurement and modeling of an ultra-wide bandwidth indoor channel," *IEEE Trans. Communications*, vol. 52, no. 10, pp. 1786-1796, Nov. 2004.

[37] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Annals of Statistics*, vol. 10, no. 4, pp. 1195-1208, Dec. 1982.

[38] B. Aksasse, L. Badidi, and L. Radouane, "A rank test approach to order estimation. Part I: 2-D AR models application," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 2069-2071, Aug. 1999.

[39] J. K. Tugnait, "Identification of linear stochastic systems via second- and fourth-order cumulant matching," *IEEE Trans. Inform. Theory*, vol. 33, no. 3, pp. 393-407, May 1987.

[40] G. B. Giannakis and S. Shamsunder, "Information theoretic criteria for non-Gaussian ARMA order determination and parameter estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 196-199, April 1993.

[41] X. Zhang and Y. Zhang, "Singular value decomposition-based MA order determination of non-Gaussian ARMA models," *IEEE Trans. on Signal Processing*, vol. 41, no. 8, pp. 2657-2664, Aug. 1993.

[42] J. K. Tugnait, "Fitting MA models to linear non-Gaussian random fields using higher order cumulants," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 1045-1050, Apr. 1997.

[43] B. Aksasse, Y. Stitou, Y. Berthoumieu, and M. Najim, "3-D AR model order selection via rank test procedure," *IEEE Trans. on Signal Processing*, vol. 54, no. 7, pp. 2672-2677, June 2006.

[44] D. Kouame and J. M. Girault, "Multidimensional multiple-order complex parametric model identification," *IEEE Trans. on Signal Processing*, vol. 56, no. 10, pp. 4574-4582, July 2008.

[45] S. Rital, A. Meziane, M. Rziza, and D. Aboutajdine, "Two-dimensional non-Gaussian autoregressive model order determination," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 426-428, Dec. 2002.

[46] J. K. Tugnait, "Linear model validation and order selection using higher-order statistics," *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1728-1736, July 1994.

[47] C. Xiao, X. Zhang, and Y. Li, "A new method for AR order determination of an ARMA process," *IEEE Trans. Signal Processing*, vol. 44, no. 11, pp. 2900-2903, Nov. 1996.

[48] G. B. Giannakis and J. M. Mendel, "Cumulant-based order determination of non-Gaussian ARMA models," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 38, no. 8, pp. 1411-1423, Aug. 1990.

[49] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 27-35, Jan. 1990.

[50] C. E. Shannon, *A mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

[51] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, pp. 37-52, 1967.

[52] M. Mishali and Y. C. Eldar, "From theory to practice: sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375-391, Apr. 2010.

[53] A. J. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.

[54] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: capacity of sampled Gaussian channels," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4889-4914, Aug. 2013.

[55] G. Ganesan and P. Stoica, "Space-time block codes: a maximum SNR approach," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1650-1656, May 2001.

[56] E. Larsson, P. Stoica, and J. Li, "Orthogonal space-time block codes: maximum likelihood detection for unknown channels and unstructured intereferences," *IEEE Trans. on Signal Processing*, vol. 51, no. 2, pp. 362-372, 2003.

[57] W. Ma, B. Vo, and P. Ching, "Blind ML detection of orthogonal space-time block codes: efficient high-performance implementations," *IEEE Trans. on Signal Processing*, vol. 54, no. 2, pp. 738-751, 2006.

[58] Y. Li, C. Georghiades, and G. Huang, "Iterative maximum likelihood sequence estimation for space-time coded systems," *IEEE Trans. on Communications*, vol. 49, no. 6, pp. 948-951, 2001.

[59] N. Ammar and Z. Ding, "Blind channel identifiability for generic linear space-time block codes," *IEEE Trans. on Signal Processing*, vol. 55, no. 1, pp. 202-217, 2007.

[60] S. Shahbazpanahi, A. Gershman, and J. Manton, "Closed form blind MIMO channel estimation for othogonal space-time codes," *IEEE Trans. on Signal Processing*, vol. 53, no. 12, pp. 4506-4517, 2005.

[61] J. Via and I. Santamaria, "Correlation matching approaches for blind OSTBC channel estimation," *IEEE Trans. on Signal Processing*, vol. 56, no. 12, pp. 5950-5961, 2008.

[62] Q. Wu and Q. Liang, "Co-prime Sampling for Higher-Order Statistics with Application to LTE Channel Estimation," *IEEE Conf. on Comm. (ICC)*, Syndney, Australia, June, 2014.

[63] V. Choqueuse, A. Mansour, and K. Yao, "Blind channel estimation for STBC systems using higher-order statistics," *IEEE Trans. on Wireless Comm.*, vol. 10, no. 2 , pp. 495-505, 2011.

[64] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. on Comm.*, vol. 46, no. 7, pp. 902-915, Aug. 1998.

[65] V. Mignone and A. Morello. "CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. on Comm.*, vol. 44, no. 9, pp. 1144-1151, Aug. 1996.

[66] W. A. Gardner, "A new method of channel identification," *IEEE Trans. on Communications*, vol. 39, no. 6, pp. 813-817, July 1991.

[67] G. B. Giannakis and M. K. Tsatsanis, "Signal detection and classification using matched filtering and higher order statistics," *IEEE Trans. Acous. Speech and Signal Proc.*, vol. 38, no. 7, pp. 1284-1296, July 1990.

[68] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," *Blind deconvolution*, Prentice Hall, pp. 8-59, 1994.

[69] Q. Wu and Q. Liang, "Co-prime sampling for higher-order statistics with application to LTE channel estimation," *IEEE International Conf. on Comm. (ICC)*, Sydney, Australia, June 2014.

[70] Q. Wu and Q. Liang, "Co-prime sampling-based third-order cumulants for UWB channel order determination," *IEEE International Conf. on Comm. (ICC)*, Sydney, Australia, June 2014.

[71] M. Ganguli, "A note on nested sampling," *Sankhy: The Indian Journal of Stat.*, pp. 449-452, 1941.

[72] F. Feroz and M. P. Hobson. "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses," *Monthly Notices of the Royal Astronomical Society*, vol. 384, no. 2, pp. 449-463, 2008.

[73] P. Pal and P. P. Vaidyanathan, "Nested arrays in two dimensions, part I: Geometrical considerations," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4694-4705, June 2012.

[74] —-, "Nested arrays in two dimensions, Part II: Application in two dimensional array processing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4706-4718, June 2012.

[75] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M.C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497-508, Apr. 2012.

[76] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Information Theory*, vol. 59, no. 9, pp. 5213-5226, Sep. 2013.

[77] H. Dai, A. F. Molisch, and H. V. Poor, "Downlink capacity of interference-limited MIMO systems with joint detection," *IEEE Trans. Wireless Communications*, vol. 3, no. 2, pp. 442-453, Mar. 2004.

[78] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Information Theory*, vol. 40, no. 6, pp. 1713-1727, Nov. 1994.

[79] O. Somekh and S. Shamai, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel with fading," *IEEE Trans. Information Theory*, vol. 46, no. 4, pp. 1401-1425, Jul. 2000.

[80] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, p. 50, Jun. 2002.

[81] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 361-374, Feb. 2006.

[82] M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, no. 3, vol. 29, pp. 439-441, May 1983.

[83] G. Caire and S. Shamai, "On the achievable throughput of a multiple-antenna Gaussian broadcast channel," *IEEE Trans. Information Theory*, vol. 49, no. 7, pp. 1691-1706, Jul. 2003.

[84] M. Karakayali , G. Foschini and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Trans. Wireless Communications*, vol. 13, no. 4, pp. 56-61, Aug. 2006.

[85] O. Somekh, B. M. Zaidel, and S. Shamai, "Sum rate characterization of joint multiple cell-site processing," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4473-4497, Dec. 2007.

[86] O. Simeone, O. Somekh, and S. Shamai, "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans. Info. Theory*, vol. 55, no. 1, pp. 190-204, Jan. 2009.

[87] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 1, pp. 8-15, Jan. 1989.

[88] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Information Theory*, vol. 18, no. 6, pp. 725-730, Nov. 1972.

[89] V. Sergio, "spectrum efficiency in the wideband regime," *IEEE Trans. Information Theory*, vol. 48, no. 6, pp. 1319-1343, Jun. 2002.

[90] R. Roy and T. Kailath,"ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 7, pp. 984-995, Jul. 1989.

[91] P, Gupta and P. R. Kumar. "The capacity of wireless networks," *IEEE Trans. Information Theory*, vol. 46, no. 2, pp. 388-404, Aug. 2000.

[92] O. Dousse, P. Thiran, and M. Hasler. "Connectivity in ad-hoc and hybrid networks," *Proc. of IEEE INFOCOM*, vol. 2, pp. 1079-1088. Jun. 2002.

[93] D. Gesbert, S. Hanly, and H. Huang, et al., "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. on Sel. Areas in Comm.*, vol.28, no.9, pp. 1380-1408, Dec. 2010.

[94] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Comm.*, vol. 6, no. 1, pp. 69-73, Jan. 2007.

[95] R. Heath, T. Wu, Y. H. Kwon, and A. Soong, "Multiuser MIMO in distributed antenna systems with out-of-cell interference," *IEEE Trans. Signal Process.*, vol. 59, No. 10, pp. 4885-4899, Oct. 2011.

[96] M. Sawahashi, Y. Kishiyama, and A. Morimoto, et al., "Coordinated multipoint transmission/reception techniques for LTE-advanced," *IEEE Wireless Comm.*, vol. 17, no. 3, pp. 26-34, Jun. 2010.

[97] W. Feng, Y. Wang, and N. Ge, et al. "Virtual MIMO in multi-cell distributed antenna systems: coordinated transmissions with large-scale CSIT," *IEEE J. Sel. Areas Commun.* Vol. 31, no. 10, pp. 2067-2081, Oct. 2013.

[98] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol.46, no.2, pp: 388-404, Mar. 2000.

[99] X. Wang and Q. Liang, "On the throughput capacity and performance analysis of hybrid wireless networks over fading channels," *IEEE Trans. Wireless Comm.*, vol. 12, no. 6, pp. 2930-2940, Jun. 2013.

[100] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[101] J. G. Proaksi and M. Salehi, *Digital Communication*, 5th edition. McGraw-Hill, 2008.

[102] S. U. Pillai and B.H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 1, pp. 8-15, Jan. 1989.

BIOGRAPHICAL STATEMENT

Qiong Wu earned the Ph.D. degree in Electrical Engineering at the Department of Electrical Engineering, University of Texas, Arlington. He received the B.S. degree in Electrical Engineering from Harbin Institute of Technology, Harbin, China, in 2006 and M.S. degree of Computer Engineering from University of Texas, Austin, Texas, USA, in 2011. His research interests include wireless communications, compressive sensing, statistical signal processing, and data mining.