Analysis of data-aided channel tracking for hybrid massive MIMO systems in millimeter wave communications

by

© *Xiaolin Pang*, BEng

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Engineering

Faculty of Engineering & Applied Science

Memorial University of Newfoundland

St. John's, NL.

March 2022

Abstract

As the data traffic in future wireless communications will explosively grow up to 1000 folds by the deployment of 5G, several technologies are emerging to satisfy this demand, including massive multiple-input multiple-output (MIMO), millimeter wave (mmWave) communications, Non-Orthogonal Multiple Access (NOMA), etc. The combination of millimeter wave communication and massive MIMO is a promising solution since it can provide tens of GHz bandwidth by fundamentally exploring higher unoccupied spectrum resources. As the wavelength of higher frequency shrinks, it is possible to design more compact antenna array with a very large number of antennas. However, this will cause enormous hardware cost, energy consumption and computation complexity of decent RF (Radio Frequency) chains. To this end, spatial sparsity is widely explored to enable hybrid mmWave massive MIMO systems with limited RF chains to achieve high spectral and energy efficiency.

On the other hand, channel estimation problem for systems with limited RF chains is quite challenging due to the unaffordable overhead. To be specific, the conventional pilot-based channel estimation requires to repeatedly transmit the same pilot because only a limited number of antennas will be activated for each time slot. Therefore, it consumes a huge amount of temporal and spectral resources. To overcome this problem, channel estimation for mmWave massive MIMO systems is still an on-going research area. Among plenty of candidates, channel tracking is the most promising one. To achieve the extremely low cost and complexity, which is also the greatest motivation of this thesis, data-aided channel tracking method is thoroughly investigated with closed-form CRLB (Cramér-Rao lower bound). In this thesis, data-aided channel tracking systems with different types of antenna, including ULA (Uniform Linear Antenna array), DLA (Discrete Lens Antenna ar-

ray) and UPA (Uniform Planar Antenna array), are comprehensively studied and proposed, and the closed-form expressions of the corresponding CRLBs are carefully derived. The numerical results of the simulations for each case are shown respectively, and they reveal that the performance of the proposed data-aided channel tracking system approaches the CRLB very well.

In addition, to further explore the data-aided channel tracking system, the multi-user scenario is investigated in this thesis. This is motivated by the highway and high-speed railway application, where overtaking operation happens frequently. In this case, the users in the same beam suffer from high channel interference, thus degrading the channel estimation performance or even causing outage. To deal with this issue, we proposed an estimated SER (Symbol Error Rate) metric to indicate if a scheduling operation is necessary to be taken place and restart of the whole channel tracking system is required. This metric is included as the *Update* phase in the proposed channel tracking method for multi-user scenario with DLA. The theoretical SER closed-form expression is also derived for multi-user data detection. The numerical results of the simulations verified the theoretical SER expression, and the scheduling metric based on the estimated SER performance is also discussed.

Acknowledgements

It is a long and productive journey for me to be here to approach graduation. Not compared with others, but only compared with myself. As I decided to start working on this area, I have spent much longer time and more effort to make up the gap. I can not forget those so many struggling nights and can not believe that I am finally here. At this moment, I would like to acknowledge and give my warmest thanks to the help and support from my university, friends and family.

Foremost, I would like to express my sincere gratitude to my supervisor Dr. Cheng Li and co-supervisor Dr. Zijun Gong for their continuous advise and support through my whole master program. Thanks for their enough patience to allow me to build solid background in this area, and also thanks for their open eyes to guide me explore widely in both research and life. Besides, I appreciate so many teaching and working opportunities, given by Dr. Li, to enrich my experience in both career and life.

Also, my sincere thank goes to Dr. Yuanzhu Chen who has moved to Queen's University. Dr. Li and he introduced me to join the WineMocol group, where I browsed and learned a lot about network area. Dr. Chen also encouraged me to sharp my skills from different perspectives, such as English practicing and sport exercises. Here, I would also thank all the WineMocol group members I have met, including Chen Zhang, Sipan Ye, Zhendong Sha, Ali Alsosool, Ali Farrokhtala and many friends whose names are not mentioned. Thank you all for letting me enjoy the inspiring discussions.

I am also grateful to my colleagues and friends I have met here, including Zijun Gong, Fan Jiang, Yuhui Song, Xiaohang Zhao, Ruoyu Su, Yi Zhang and many friends whose names are not mentioned. Among them, I would especially thank Dr. Fan Jiang and Yuhui

Song for their valuable guidance and advise at the beginning of this journey. I can not imagine how I could make it through all these years without their support and accompany.

Last but not the least, a big thank must go to my families for their selfless love, support and caring for educating and preparing me for my future. Even though we have to be separated at the opposite sides of the earth and there are over 12 hours time difference, our love makes us together, and provides me endless motivation and power to be here.

— Xiaolin Pang

Contents

List of Figures

List of Notations and Abbreviations

5G	Fifth Generation
MIMO	multiple-input multiple-output
mmWave	millimeter wave
NOMA	non-orthogonal multiple access
RF	radio frequency
CRLB	Cramér-Rao lower bound
ULA	uniform Linear Array
DLA	discrete Lens Array
UPA	uniform Planar Array
EM	electromagnetic
GSM	global system for mobile communications
GPRS	general packet radio service
QoS	quality of service
IoE	internet of everything
BS	base station
MMSE	minimum mean-square error
MF	matched filter
LoS	line-of-sight
NLoS	non-line-of-sight
SINR	signal to interference-plus-noise ratio
SNR	signal to noise ratio

BER	bit error rate
SER	symbol error rate
ZF	zero forcing
DFT	discrete Fourier transform
FFT	fast Fourier transform
IDFT	inverse-DFT
IFFT	inverse-FFT
IUI	inter-user interference
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
SDMA	space division multiple access
PLL	phase-locked loop
CSI	channel state information
CSCG	circularly symmetric complex Gaussian
DoA	direction of arrival
ML	maximum likelihood
MSE	mean-square error
NMSE	normalized mean-square error
AWGN	additive white Gaussian noise
QPSK	quadrature phase shift keying
EVD	eigenvalue decomposition
TDD	time division duplex

LS	least square
PDF	probability density function
$(\cdot)^T$	matrix transpose
$(\cdot)^H$	matrix conjugate transpose
$(\cdot)^*$	element-wise conjugate
$(\cdot)^{-1}$	matrix inversion
$\lfloor \cdot \rfloor$	the floor function
$E\{\cdot\}$	mathematical expectation
$\exp(\cdot)$	exponential function
$\ln(\cdot)$	natural logarithmic function
$\mathbf{I}_L$	the size L identity matrix
$\mathrm{tr}\{\mathbf{A}\}$	the trace of matrix $\mathbf{A}$
$\mathrm{diag}\{\mathbf{A}\}$	a column vector with diagonal elements in $\mathbf{A}$
$\mathrm{diag}\{\mathbf{v}\}$	a diagonal matrix with diagonal elements in $\mathbf{v}$

Chapter 1

Introduction

1.1 Introduction

In modern wireless communication, the way of transmitting and receiving signals is based on the EM (electromagnetic) wave in the air, and the one consisting of radio telephones and mobile network is called cellular network. Cellular network, as an indispensable part of wireless communication, has evolved from 1G to 5G in the past four decades, and will enter 6G era in the near future, which shows an exponential growth in the field of wireless communication over years.

The first generation (1G) cellular network was launched in 1980s, which is called Nordic Mobile Telephone (NMT), only providing voice service operating at 800 - 900 MHz frequencies with 30 KHz bandwidth. To further provide data service, the second generation (2G) cellular technology was introduced in 1990s, which is based on digital system to provide voice and data service simultaneously. The well-known GSM (Global System for Mobile)

system is the first 2G network, and other variant of 2G technologies are GPRS (General Packet Radio Service), which is used to access WAP /internet services to the customers, and EDGE (Enhanced Data rate for GSM Evolution), which is also known as 2.75G to provide maximum data rate of 384 Kbps. In 2G era, circuit switching network is used for voice and packet transmission form source to destination, and digital encryption is also first introduced in cellular networks. At the beginning of 2000s, the third generation (3G) of mobile communication system was introduced to improve voice services, data throughput, high QoS (Quality of Service) and information security. In 3G era, 144 Kbps data rate for mobile user, 384 Kbps for pedestrian user and 2 Mbps for indoor user were satisfied. The key technique for 3G is CDMA (Code Division Multiple Access), and packet switching is used for voice and data communication. As for the past decade, the fourth generation (4G) was first launched in 2010 with several significant improvements to its predecessors, such as 40 MHz capacity, 100 Mbps peak speed during handoff stages, all IP network used as switching type and internet-based core network. It is obvious that the major advancement of 4G over 3G is its higher bandwidth and data throughout, which is typically from 100 Mbps to 1 Gbps. Even though MIMO (Multiple Input Multiple Output) is already employed in 4G, it is further exploited as massive MIMO in fifth generation (5G) cellular network, which has started being deployed in 2020. As one important task of 5G is to support fast and reliable communication features, such as ultra HD video streaming, mobile full HD TV, Augmented Reality (AR), holographic communication, etc., high data throughput is still the biggest challenge in 5G. The 5G cellular network promptly states that the requirement is at least 1 Gbps or more data rate to support virtual reality applications along with up to 10 Gbps data speed to support mobile cloud service. To achieve this, massive MIMO

and millimeter wave techniques are both key features for 5G successful deployment with high capacity.

As the deployment of 5G is the first step for people to approach a new world in terms of Internet of Everything (IoE), data traffic will explosively grow up to 1000 folds in future communications. To meet this demand, massive multiple-input-multiple-output (MIMO) in sub-6GHz frequency was proposed in 2010 to provide great gain by using the large antenna arrays at BS (Base Station) compared with the number of users in the same cell [3]. Even with low-complexity algorithms, such as Minimum Mean Square Error (MMSE) estimator for channel estimation and Matched Filter (MF) for data detection, massive MIMO can still improve energy and spectral efficiencies to a great extent [4]. Since the near-optimal performance of the traditional massive MIMO in sub-6GHz is getting hard to satisfy the data demand for future communications, more attention is concentrated on millimeter wave (mmWave) spectrum (30 - 300 GHz), which is able to provide up to several GHz bandwidth by exploiting higher unoccupied frequency resources rather than improving the spectrum efficiency with current scarce bandwidth [5]. However, mmWave communication suffers from the severe signal attenuation induced by such high carrier frequency even for Line-of-Sight (LoS) path [6]. To this end, it turns out to be a natural marriage between massive MIMO and mmWave communication to compensate for order-of-magnitude path loss. As the wavelength shrinks with higher frequency, a more compact transceiver design becomes possible to benefit massive MIMO architecture, which consists of at least hundreds of antennas, thus improving the power gain. Besides, beamforming is another key technology in mmWave massive MIMO systems to compensate path loss with directional transmission [7].

In communications, beamforming is usually paired with combining, where data is sent with different weights as coefficients and coherently combined using a counterpart receiver to maximize the output SNR (Signal to Noise Ratio). The gain in the resulting SNR is in terms of increasing degree of freedom, which can be classified into two parts, multiplexing gain and diversity gain [8]. Specifically, multiplexing gain contributes to the higher achievable rate and diversity gain, explicitly revealed in the improved SNR, results in the better BER (Bit Error Rate) or SER (Symbol Error Rate) performance. This increasing degree of freedom is in terms of beamforming gain at the transmitter side, whereas it is called as combining gain at the receiver side.

Taking the transmitter side as an example, it is widely accepted to classify the beamforming as digital beamforming and analog beamforming [7]. Analog beamforming consists of a group of phase shifters which are dealing with the phase change of the RF (Radio Frequency) band signals, so the size of analog beamformer depends on the number of antennas. The popular types include phase-shifters, switches and antenna lens, etc. On the other hand, digital beamforming is equivalent to precoder which is dealing with both phase and power of the digital baseband signals, thus the size of digital beamformer depends on the number of RF chains. Since digital beamforming is a type of signal processing technique, some mature algorithms are considered such as Zero Forcing (ZF) and MMSE. To sum up, the general conclusion is that analog beamforming is efficient whereas digital beamforming is more flexible [9].

However, by adapting beamforming with mmWave communication, the conventional digital beamforming causes unaffordable power consumption and hardware cost since it requires one decent RF chain for each antenna, where the number of antennas is quite

large in mmWave communication. This leads to opt for hybrid beamforming operating in both baseband and RF domains jointly, which is still an ongoing research area [9]. Besides, another motivation for hybrid beamforming is the sparsity of mmWave channels with limited RF chains. To this end, the main idea of hybrid beamforming is that analog beamforming is still performed at the RF end, while the size of digital precoder can be dramatically reduced with a small number of RF chains, which also enables the systems with limited RF chains.

As a result, hybrid beamforming has gained a lot of interest for various mmWave communication systems with limited RF chains in the past decade. At the meantime, combining at the receiver side is an exact similar counterpart following the same protocol. One of these promising solutions for mmWave communication is Discrete Lens Antenna (DLA-based) system, which is reported in [10] with experimental result. In [10], the authors conducted the experimental measurements with a prototype system and demonstrated the utility of this emerging array form. Furthermore, [2] mathematically proved that DLA array serves as a 'sinc' function, or equivalently the Discrete Fourier Transform (DFT) operation of the original array response of a Uniform Linear Array (ULA). In other words, most energy is concentrated on a directional beam, leading to the fact that DLA array acts as an analog beamformer itself via the continuous aperture lens antenna array which is controlled by the beam selector. Specifically, most energy will be only concentrated on two antennas [1]. By applying DLA, the cost for analog beamforming is eliminated and the complexity will be dramatically reduced. [2] also compared DLA-based system performance with the conventional ULA-based mmWave communication systems and showed that this beamspace MIMO with DLA achieved similar spectral efficiency with significantly less RF

chains and lower processing complexity. However, the drawback is that the performance of this analog beamformer is fixed due to its optical characteristics, where dynamical adaption is not allowed such as spatial rotation. Moreover, the channel estimation remains as a challenge in mmWave massive MIMO systems with limited RF chains due to high resource consumption, which is the key motivation of this thesis. In the following part of this chapter, we will briefly introduce common characteristics of the wireless channel, massive MIMO and mmWave communication separately, and then elaborate the challenge in this area and motivation of this thesis. The last part of this chapter is the thesis organization. Throughout this thesis, we should note here that Doppler effect estimation is excluded by assuming the Doppler effect is perfectly eliminated via PLL (Phase-Locked Loop) at the RF end.

1.2 Challenges and motivation

To integrate these emerging techniques we have introduced in the previous sections, mmWave massive MIMO systems make our highest priority since the combination of both can provide high spectral and energy efficiency. Furthermore, mmWave massive MIMO systems with limited RF chains heave into our sight because they can dramatically reduce the hardware cost and computational complexity at the same time via hybrid beamforming/combining. Specifically, DLA -based mmWave massive MIMO system with limited RF chains firstly attracts our attention due to its promising prospect. However, real -time channel estimation is quite challenging for any system with limited RF chains, especially in high mobility scenario.

Let us take the highway application as an example. Small BSs are distributed as RSUs (Road Side Units), and devices on the vehicles are served as users in cellular network. In this scenario, LoS path generally dominates, which is natural to be adaptive with mmWave communication. Moreover, the severe signal attenuation will be compensated by concentrating energy on a narrow beam via massive MIMO architecture with very large number of antenna elements. To jointly reduce the cost and complexity, DLA-based mmWave massive MIMO system is a feasible solution for this highway application, which can be extended as general mmWave massive MIMO systems with limited RF chains. However, channel estimation problem in this case is quite challenging due to the unaffordable channel estimation overhead. To be specific, the pilot overhead is inevitable by applying conventional real-time pilot training schemes, where pilot-based channel estimation needs to be conducted for every time slot. Besides, due to the constraint of limited RF chains, only a small portion of antenna elements are activated for each time slot, thus requiring repeated transmission of the same pilot, which consumes a huge amount of resource. To overcome this problem, channel estimation for mmWave massive MIMO systems is still an on-going research area. Channel tracking method, which highly depends on the temporal correlation between adjacent time slots, is a promising candidate instead of conventional channel estimation for each time slot. This is the greatest motivation of this thesis to investigate data-aided channel tracking method for mmWave massive MIMO systems.

By taking multi-user scenario into account, channel tracking is also challenging due to inter-user interference (IUI). For data aided channel tracking method, the estimation accuracy is highly depending on the data detection performance, which is quite vulnerable to the IUI. However, due to the spatial resolution degradation issue in DLA-based systems,

IUI issue is inevitable for successful channel tracking. This is the primary motivation of the forth chapter of this thesis. Meanwhile, it is instructive to investigate different channel tracking methods with different antenna array forms, and individually evaluate their system performance, which constitutes my major work and will be present in the following chapters of this thesis. To enhance the background and comprehensive understanding of the whole system, we start with data-aided channel tracking scheme for mmWave massive MIMO system with conventional ULA, and thoroughly analyze the estimation accuracy with Cramér Rao Lower Bound (CRLB) derivation, which is presented in the third chapter. In addition, we further extend the data-aided channel tracking method to the mmWave massive MIMO systems with UPA (Uniform Planar Array), which is elaborated in the fifth chapter associated with its accuracy analysis.

1.3 Thesis organization

First of all, the basic framework of data-aided channel tracking is learned from [1], which is applied for DLA-based mmWave communication system with limited RF chains for single user scenario. However, the closed-form expression of CRLB is missing in [1]. Therefore, in this thesis, we comprehensively studied channel tracking schemes with different antenna types, gave the missing closed-form CRLB in [1], and extended the work for multi-user scenario. The organization of the entire thesis is as follows.

In Chapter 2, the fundamentals of wireless channel characteristics, massive MIMO and millimeter wave communications are introduced and the literature review of some existing work is followed. In Chapter 3, the ULA-based data-aided channel tracking method for

mmWave massive MIMO system is presented and the corresponding CRLB with respect to spatial signature estimation and DoA estimation are both derived respectively. In Chapter 4, the DLA-based data-aided channel tracking method for mmWave massive MIMO system is studied, and the closed-form CRLB is carefully derived. In this chapter, we replay the results in [1] and verifies the theoretical CRLB analysis. In Chapter 5, the extensional work for multi-user scenario is explored, which mainly deals with the co-channel issue. In this chapter, the joint channel truncating algorithm is proposed and the predicted SER based scheduling metric is achieved, which is supported by the theoretical results. In Chapter 6, the UPA-based data-aided channel tracking method for mmWave massive MIMO system is presented and the corresponding CRLB is derived, which is validated by the simulation results. The last chapter summaries the entire thesis and talks about the possible work we can do in the future.

Chapter 2

Literature review

2.1 Backgrounds

2.1.1 Characteristics of wireless channel

In wireless communications, signals are transmitted through the medium as electromagnetic waves between the transmitter and the receiver. The received signals may be quite different from the source due to reflection, refraction, scattering and absorption. In other words, because of the inherent characteristics in the propagation environment, the received signal will be the superposition of shifted and squeezed replicas of the source, resulting in either constructive or destructive interference of the channel envelop, which is also referred to as *fading effect* [11]. Generally, *fading* can be classified as large-scale fading and small-scale fading, which will be illustrated.

The major reasons causing *fading effect* include *multipath* propagation and *mobility*. To describe these characteristics, suppose the transmitter and receiver are placed in free

space and there is a fixed wall introducing a reflected path, as depicted in Figure 2.1, where the direct path is indicated as orange and the reflected path is indicated as green.

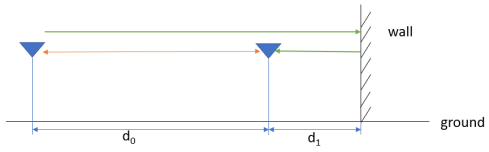

Figure 2.1: Reflecting wall and fixed antenna scenario.

If a sinusoid wave is transmitted at carrier frequency f_c, then the received signal via the direct path will be $r_0(t) = \beta_0 \cos(2\pi f_c t - \theta_0)$, and the received replica via the reflected path will be $r_1(t) = \beta_1 \cos(2\pi f_c t - \theta_1)$, where β_0 and β_1 are the attenuation factors, and two shifted phases are denoted as $\theta_0 = 2\pi f_c d_0/c$, and $\theta_1 = 2\pi f_c(d_0 + 2d_1)/c$, where c is the speed of light. Therefore, the received signal is the superposition of the two

$$\begin{aligned} r(t) &= \beta_0 \cos(2\pi f_c t - \theta_0) + \beta_1 \cos(2\pi f_c t - \theta_1) \\ &= (\beta_0 \cos\theta_0 + \beta_1 \cos\theta_1)\cos(2\pi f_c t) + (\beta_0 \sin\theta_0 + \beta_1 \sin\theta_1)\sin 2\pi f_c t. \end{aligned} \tag{2.1}$$

Let us define $\tilde{\beta}_0 = \beta_0 \cos\theta_0 + \beta_1 \cos\theta_1$, $\tilde{\beta}_1 = \beta_0 \sin\theta_0 + \beta_1 \sin\theta_1$ and there must be a α satisfying $\cos\alpha = \tilde{\beta}_0/\sqrt{\tilde{\beta}_0^2 + \tilde{\beta}_1^2}$ and $\sin\alpha = \tilde{\beta}_1/\sqrt{\tilde{\beta}_0^2 + \tilde{\beta}_1^2}$, then we have

$$r(t) = \sqrt{\tilde{\beta}_0^2 + \tilde{\beta}_1^2}\cos(2\pi f_c t - \alpha), \tag{2.2}$$

which indicates the resulting strength of the received signal $A = \tilde{\beta}_0^2 + \tilde{\beta}_1^2$ will be

$$\begin{aligned} A &= (\beta_0 \cos\theta_0 + \beta_1 \cos\theta_1)^2 + (\beta_0 \sin\theta_0 + \beta_1 \sin\theta_1)^2 \\ &= \beta_0^2 + \beta_1^2 + 2\beta_0\beta_1 \cos(\theta_0 - \theta_1) \\ &= (\beta_0 - \beta_1)^2 + 4\beta_0\beta_1 \cos^2\left(2\pi f_c \frac{2d_1}{c}\right). \end{aligned} \tag{2.3}$$

At this point, we can clearly see that the signal strength is a function of the frequency f_c. Specifically, the signal strength is a periodic function of f_c with the period of $\frac{c}{2d_1}$, which is actually the reciprocal of the delay spread between the direct path and reflected path. This characteristic introduces the constructive and destructive interference pattern upon the signal strength depending on frequency f_c, thus causing frequency selective fading for the large frequency band and flat fading for the small frequency band.

Furthermore, taking mobility into consideration, if the receiver is moving away from the reflected wall at the speed of v and the propagation delay is constantly varying, which is shown in Figure 2.2, then the wireless channel is no longer an LTI (Linear Time-Invariant) system. If we still transmit a sinusoid wave at frequency f_c, the received signals will still be represented as $r_0(t) = \beta_0(t)\cos(2\pi f_c t - \theta_0)$ and $r_1(t) = \beta_1(t)\cos(2\pi f_c t - \theta_1)$, but with the attenuation factors and phase shift become a function of time t as $\theta_0(t) = 2\pi f_c(d_0 - vt)/c$ and $\theta_1(t) = 2\pi f_c(d_0 + 2d_1 + vt)/c$.

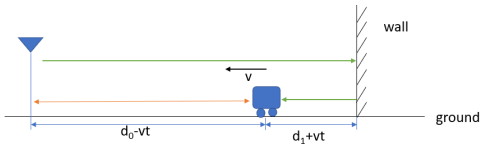

Figure 2.2: Reflecting wall and moving antenna scenario.

Similar to the previous derivation, the received signal is also a superposition of the direct path and the reflected path. Also define $\tilde{\beta}_0(t) = \beta_0(t)\cos\theta_0(t) + \beta_1(t)\cos\theta_1(t)$, $\tilde{\beta}_1(t) = \beta_0(t)\sin\theta_0(t) + \beta_1(t)\sin\theta_1(t)$ and there must be a α satisfying $\cos\alpha = \tilde{\beta}_0(t)/\sqrt{\tilde{\beta}_0^2(t) + \tilde{\beta}_1^2(t)}$ and $\sin\alpha = \tilde{\beta}_1(t)/\sqrt{\tilde{\beta}_0^2(t) + \tilde{\beta}_1^2(t)}$, then the total signal strength is

$$\begin{aligned} A &= \tilde{\beta}_0^2(t) + \tilde{\beta}_1^2(t) \\ &= \left(\beta_0(t)\cos\left(\theta_0(t)\right) + \beta_1(t)\cos\left(\theta_1(t)\right)\right)^2 + \left(\beta_0(t)\sin\left(\theta_0(t)\right) + \beta_1(t)\sin\left(\theta_1(t)\right)\right)^2 \\ &= \beta_0^2(t) + \beta_1^2(t) + 2\beta_0(t)\beta_1(t)\cos\left(\theta_1(t) - \theta_0(t)\right) \\ &= \left(\beta_0(t) - \beta_1(t)\right)^2 + 4\beta_0(t)\beta_1(t)\cos^2\left(2\pi f_c(d_1 + vt)/c\right). \end{aligned} \tag{2.4}$$

From this, we can clearly see that the signal strength is not only a function of carrier frequency f_c but also a periodic function of time t with the period of $\lambda/2v$, which depends on the velocity v. In other words, the signal strength oscillates by every $T = \lambda/4v$ falling from the peak to the valley, which means the channel is constantly varying with time at fast speed. This is referred as Doppler spread to describe how fast the channel varies, thus causing fast fading with high velocity and slow fading with low velocity.

To sum up, the fading effect can be classified as large-scale fading and small-scale fading based on distance between the transmitter and the receiver. In Equation (2.4), there are three components affecting the signal strength: $\beta_0(t)$, $\beta_1(t)$, and the cosine function depending on the carrier frequency f_c and the moving velocity v. When the receiver moves for a short distance, $\beta_0(t)$ and $\beta_1(t)$ change non-significantly, which is called attenuation factors. In practical scenario, there is possibility of suffering propagation blockage due to large obstacles. This is referred to as shadowing effect, which is commonly modeled as log-

normal distribution. Therefore, the *large-scale fading* contains attenuation and shadowing effect.

On the other hand, the cosine function in (2.4) is dominated by the carrier frequency and the velocity, and causes constructive and destructive interference on signal envelope even over a very short distance, which is referred to as *small-scale fading*. Generally, *coherence bandwidth* is the concept to describe how fast the fading coefficient changes with the carrier frequency. If the delay spread between two frequencies is larger than coherence bandwidth, their fading coefficients differ significantly, which is referred to as *frequency-selective fading*. Otherwise, it is referred to *flat fading*. Besides, *coherence time* is the concept to describe how fast the channel changes with time. If the Doppler spread of two paths is larger than coherence time, it suffers from *fast fading*, which means the channel is time-variant. Otherwise, it is referred to *slow fading*, which means the wireless channel is still viewed as a LTI system. In this thesis, these terms will be frequently appeared. In most cases, we assume flat fading with the help of Orthogonal Frequency Division Multiplexing (OFDM) and slow fading even in the high-mobility scenario supposing that the Doppler effect has been compensated at the RF end by Phase-Locked Loops (PLLs).

2.1.2 Introduction to massive MIMO

In traditional wireless communication systems, researchers concentrate on exploiting the diversity in one of three transformation domains, which are time domain, frequency domain and code domain. During the first decade of 21-th century, numerous works have been proposed regarding multi-user multiple-input multiple-output (MU-MIMO) communication system by investigating the spatial degree of freedom for the trade-off between

diversity and multiplexing [12–14]. However, Orthogonal Frequency Division Multiple Access (OFDMA) is generally mandatory with Space Division Multiple Access (SDMA) in traditional MU-MIMO systems to mitigate interference due to channel correlation, which degrades the spectral efficiency. Until 2010, massive MIMO technique was proposed to bring new hopes [3]. The key idea of massive MIMO is to increase the number of antennas without bound to allow each user's channel vector to be asymptotically orthogonal, where uncorrelated noise and inter-user interference become negligible [3]. The idea of massive MIMO addresses two fundamental issues in conventional wireless communication systems by improving the spectral efficiency and energy efficiency simultaneously even with simple MF receiver with perfect or imperfect channel-state information (CSI) [4]. By doing this, it is possible to increase the total capacity by an order of magnitude and fully enable SDMA to allow users to use all the time-frequency resources at the same time. In other words, OFDMA can be totally discarded in massive MIMO systems where SDMA is the better choice. To illustrate this characteristic, we start with statistical channel model with conventional point-to-point MIMO system.

Specifically, consider a narrowband time-invariant MIMO system with N_t transmit antennas and N_r receive antennas. Then the system model can be described as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.5}$$

where $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the deterministic channel matrix with each element as the complex channel gain for Rayleigh fading channel, $\mathbf{y}$ and $\mathbf{x}$ denote the transmitted and received channel vectors respectively, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_r})$ is the CSCG (Circularly Symmetric

Complex Gaussian) noise vector. If we assume full CSI, then the channel matrix can be decomposed into a set of parallel, independent scalar Gaussian sub-channels. For example, by applying singular value decomposition (SVD), the channel matrix $\mathbf{H}$ will be

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H = \sum_{k=1}^{n_{min}} \lambda_k \mathbf{u}_k \mathbf{v}_k^H, \tag{2.6}$$

where $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ are unitary matrices with $\mathbf{u}_k$, $\mathbf{v}_k$ as their k-th column respectively, $\mathbf{\Lambda} \in \mathbb{R}^{N_r \times N_t}$ is a rectangular diagonal matrix whose diagonal elements are non-negative real numbers, which are the singular values of the channel matrix with the descending order of $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{n_{min}}$, and $\lambda_{n_{min}} \leq \min(N_t, N_r)$ denotes the number of non-zero singular values, or equivalently the rank of $\mathbf{H}$. Assume both the transmitter and receiver have the full CSI, appropriate precoding and combining can be performed on both sides, such as $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{x}$ and $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$. Then the total spectral efficiency can be easily derived as

$$\eta_s = \sum_{k=1}^{m_{min}} log_2\left(1 + \frac{\lambda_k^2 P_k}{\sigma^2}\right), \tag{2.7}$$

where P_k is the power of the k-th symbol in $\mathbf{x}$. Apparently, the channel capacity will increase linearly while the numbers of antennas at the transmitter and receiver keep growing. This is because that the same time-frequency resource is repeatedly used by many channels, which is referred as to *multiplexing*. However, the total capacity is not only dependent on the number of antennas, but is also related to the rank of the channel matrix $\mathbf{H}$, which means the number of independent parallel sub-channel that can be multiplexed. Therefore, the channel correlation is always a problem to overcome in conventional MIMO system, which limits the energy and spectral efficiency. At this point, massive MIMO mathematical

model is proposed in [3] to allow the number of antennas growing without bound, which enable a key channel characteristics as

$$\lim_{N_r \to +\infty} \mathbf{H}^H \mathbf{H} \to \mathbf{I}_{N_t}. \tag{2.8}$$

That is also to say, for large number of receive antennas, each user's channel vector will be asymptotically orthogonal based on Lindeberg-Lévy central limit theorem [3], thus eliminating random noise effect and inter-user interference. Therefore, only inter-cell interference remains in the massive MIMO systems, which also leads to pilot contamination. The greatest contribution of massive MIMO is to enable all users to share the same time-frequency resources by only employing SDMA. By doing this, energy and spectral efficiency have been significantly improved.

Besides the exploitation in time-frequency domain of MIMO systems, spatial domain is also commonly exploited to gain more degrees of freedom, especially when massive MIMO and mmWave techniques are combined. In the next section, we will briefly introduce the mmWave communication system and its high spatial sparsity with spatial channel modeling.

2.1.3 Introduction to mmWave communication

As the demand of data traffic has been growing extremely fast over the past two decades, the current cellular network has been moved from 2 GHz to 5 GHz frequency band to gain more available bandwidth [11]. Unfortunately, it becomes crowded at present and hard to satisfy the heave data traffic demand. Consequently, higher frequency band at-

tracts researchers' attention to exploit higher unoccupied frequency resources, which is in terms of millimeter wave (mmWave) frequency band, ranging from 20 to 60 GHz, since its wavelength is about several millimeters [5]. However, mmWave communication suffers from severe path loss due to such high frequency, which prevents it to be implemented in practice. At this point, massive MIMO is a promising solution, while being combined with mmWave communication, to compensate the huge attenuation [15]. To be specific, massive MIMO not only provides a large amount of power gain via beamforming, but also makes it possible to fully employ SDMA based on its high spatial sparsity with limited RF chains. Next we will introduce the high spatial sparsity of massive MIMO mmWave communication system with spatial channel modeling.

To facilitate our illustration, assume Uniform Linear Antenna (ULA) array is employed at the base station (BS) with omnidirectional antenna elements, which is separated from each other with equal space, typically as half wavelength $\lambda/2$. Due to the severe path loss, the Line-of-Sight (LoS) path dominates. On the other hand, the reflected path is several times weaker than the direct signal path [5]. Assume free space without any reflector or scatters, and only LoS path between the BS and the single user with single omnidirectional antenna. For the k-th receive antenna element, the uplink continuous-time impulse response $h_k(t)$ is given as

$$h_k(t) = \alpha_k \delta\left(t - d_k/c\right), \ \ k = 1, ..., N_r \tag{2.9}$$

where a_k denotes the attenuation factor, c is the speed of light, d_k represents the distance between the transmit antenna and the k-th receive antenna element, and N_r is the number

of receive antennas. Considering $d \gg \frac{N_r \lambda}{2}$, or equivalently the distance between the transmitter and receiver is much larger than the physical dimension of the antenna array, all the path loss factors between transmitter and every antenna element are almost identical. Therefore, the equivalent baseband channel model is given as

$$h_k^{(b)}(t) \approx \alpha_k \exp\left(-\frac{j2\pi f_c d_k}{c}\right) \delta\left(t - d_k/c\right), \tag{2.10}$$

where d_0 is the distance between the transmitter and the reference antenna element of the ULA. To further simplify the channel model, if the angle of arrival (AoA) is ϕ, then the first-order approximation of d_k will be $d_k \approx d_0 + \frac{(k-1)\lambda \cos\phi}{2}$. As a result, the channel response will be approximated as

$$h_k \approx \alpha \exp\left(-\frac{j2\pi f_c \left(d_0 + \frac{(k-1)\lambda \cos\phi}{2}\right)}{c}\right) = \alpha \exp\left(\frac{-j2\pi d_0}{\lambda}\right) \exp\left(-j\pi(k-1)\cos\phi\right). \tag{2.11}$$

Let us define $\tilde{\alpha} = \alpha \exp\left(\frac{-j2\pi d_0}{\lambda}\right)$ and $\omega = \pi \cos\phi$, then the standard channel response is modeled as

$$\mathbf{h}_k = \tilde{\alpha}\mathbf{e}(\omega_k), \tag{2.12}$$

where $\mathbf{e}(\omega_k)$ denotes the steering vector of the k-th user, and ω_k is commonly called spatial

signature, which is given as

$$\mathbf{e}(\omega_k) = \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1 \\ \exp(-j\omega_k) \\ \vdots \\ \exp\left(-j(N_r-1)\omega_k\right) \end{bmatrix} \tag{2.13}$$

At this point, we can extend it to MIMO system model with LoS path, in terms of its angle of departure (AoD) and AoA, where the channel matrix will be

$$\mathbf{H} = \beta \mathbf{e}_{N_r}(\omega_r)\mathbf{e}_{N_t}^T(\omega_t), \tag{2.14}$$

where $\beta = \alpha\sqrt{N_r N_t}\exp(-j2\pi d/\lambda)$, N_t and N_r represent the number of transmit and receive antennas respectively, and $\mathbf{e}_{N_r}(\omega_r)$ and $\mathbf{e}_{N_t}(\omega_t)$ are the steering vectors on both transmitter side and receiver side, respectively. Here, it is natural to assume LoS path dominating in mmWave communication system, so it results in *sparsity* in space domain. To see this feature explicitly, we take single user with single antenna as an example, where the BS is equipped with ULA with N antenna elements, the received uplink signal will be given as

$$\mathbf{y} = x\beta\mathbf{e}(\omega) + \mathbf{n}, \tag{2.15}$$

where x is the transmitted symbol, and $\mathbf{n}$ represents the CSCG noise component. By performing M-point Inverse Discrete Fourier Transform (IDFT) of the received signal to

obtain its spatial spectrum as following

$$\mathbf{Y}_\omega = \frac{1}{\sqrt{N}} \sum_{m=1}^{M} \mathbf{y}[m] e^{jm\omega}. \tag{2.16}$$

Therefore, the envelope of each sample $|\mathbf{Y}_\omega[m]|$ is given as

$$|\mathbf{Y}_\omega[m]| = \frac{1}{N} \cdot \left| \frac{\sin \frac{N(\omega - n\omega_0)}{2}}{\sin \frac{\omega - n\omega_0}{2}} \right| = \frac{1}{N} \cdot \mathrm{sinc}(\omega - n\omega_0), \tag{2.17}$$

At this point, we can clearly see that the spatial spectrum is a sinc function, where most of the energy concentrates on the two samples in the main lobe, which is referred as to *spatial sparsity*. This critical characteristics enables mmWave massive MIMO systems with limited RF chains, and also paves the way for channel estimation and tracking with less pilot overhead, which will be illustrated in detail in the next chapter.

2.2 Literature review

As discussed in the first chapter, acquiring the accurate CSI (channel state information) in mmWave massive MIMO communication systems is still one of the main challenges due to the enormous cost, in terms of the precious time-frequency resources and hardware and energy. As the carrier frequency could rise up to Terahertz frequency, the traditional channel estimation techniques are no longer applicable. Moreover, large training resource is required since the cell capacity keeps increasing. Another trade-off between hardware cost and computational complexity arises by choosing fully digital antenna array or hybrid ones [16]. Therefore, we will review existing channel estimation methods based on these

two antenna array structures.

The case of fully digital antenna array, each antenna element on the array requires a decent RF chain to control the corresponding amplitude and phase of the transmitted signal, which introduces enormous hardware cost and energy consumption in mmWave massive MIMO systems. From the array signal processing perspective, beam alignment methods are typically applied to extract the channel information by scanning the entire space. Both [17] and [18] employ the hierarchical codebook to implement the beam alignment, whereas [18] is superior to [17] by further refining the codebook to allow the overlapped searching beams, thus reducing the overhead. [19] proposed a grid-of-beams (GoB) based method to obtain the angle information by selecting the best combinations of transmit and receive beams. However, such approaches suffer from the high computational overhead due to the exhaustive sequential search. Besides, the conventional pilot training based channel estimation can also be performed to extract CSI by consuming large amount of time-frequency resources and energy cost [8]. In addition, the fully digital antenna arrays suffer from vibration and movement of beams due to varying weather conditions, which leads to the frequent calibration overhead of the beam alignment. Therefore, the communication systems with limited RF chains becomes a growing field of interest for researchers due to its low cost and low complexity [20–23].

To enable limited RF chains, it is based on the fact that the mmWave channel is generally sparse in the spatial domain since the LoS path dominates due to its high signal attenuation and absorption. Taking the narrowband ULA array response of the k-th user

as an example, where M antennas are equally spaced by d,

$$\mathbf{h}_k = \beta e^{j\phi}\mathbf{e}[\omega_k], \tag{2.18}$$

where β is the attenuation factor, ϕ is the random phase delay and the steering vector is defined as $\mathbf{e}[\omega_k] = \left[1, e^{j\omega_k}, ..., e^{j(M-1)\omega_k}\right]^T$ with $\omega_k = \frac{2\pi d \cos\theta_k}{\lambda}$ and θ_k is the DoA (Direction of Arrival) of the incoming signal corresponding to the k-th user.

Definition 1. *The linear projection of the array response onto the Fourier orthonormal basis will be the* ***beamspace channel response*** $\mathbf{h}^B$ *for ULA*

$$\mathbf{h}_k^B = \mathbf{F}\mathbf{h}_k, \tag{2.19}$$

where $\mathbf{F}$ *is the* M*-point DFT (Discrete Fourier Transformation) matrix.*

Authors in [24] have demonstrated its sparsity where the most energy concentrates on two samples in the main lobe. To be more specific, in mmWave massive MIMO system, the received signal is given as

$$\mathbf{y} = \mathbf{H}\mathbf{x}. \tag{2.20}$$

The dimension of $\mathbf{y}$ is generally large due to the large size of the antenna array, whereas the dimension of $\mathbf{x}$ is much smaller than that of $\mathbf{y}$ based on the two facts: (1) LoS path dominates in mmWave scenario, and (2) the number of users is much smaller than that of antenna elements in massive MIMO scenario.

This spatial sparsity property makes it possible to exploit the efficiencies that can be obtained through compressed sensing, as illustrated in Theorem 1 [25], which means

the sparse representation of the original signal can be acquired from a limited number of linear projections [26]. In other words, $\mathbf{h}_k$ and $\mathbf{h}_k^B$ are different representations of the same essential signal in different domains. Once the RIP (Restricted Isometry Property) is satisfied by the measurement matrix, it is reliable to uniquely recover the original signal from a sparse solution to solve ℓ_0 norm [1] problem of an underdetermined linear system of equations [27]. To sum up, the signal recovery and sparse approximation are dual to each other [27]. The channel estimation can be seen as a sparse signal recovery problem from the compressed sensing perspective, where the dimension of the received signal is much larger than that of the transmitted signal in mmWave massive MIMO scenario.

Theorem 1. *[25] Given a sensing problem* $\mathbf{v} = \mathbf{\Phi s}$*, the real-valued discrete signal* $\mathbf{s} \in \mathbb{R}^d$ *is measured as a low-dimensional measurement* $\mathbf{v} \in \mathbb{R}^N$ *via a measurement matrix* $\mathbf{\Phi} \in \mathbb{R}^{N \times d}$*, where the measured signal is sparse, i.e.* $\|\mathbf{s}\|_0 = m \ll d$*. Any matrix* $\mathbf{\Phi}$ *satisfies RIP (Restricted Isometry Property) [28] of order* $2m$*, the* m*-sparse signal* $\mathbf{s}$ *can be uniquely recovered via a suitable algorithm to solve the following problem*

$$(\ell_0) \quad \hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_0 \ \textit{such that}\ \mathbf{v} = \mathbf{\Phi s}. \tag{2.21}$$

Generally, to recover the sparse signal, we need to find an algorithm to solve the undetermined linear system of equations, and such techniques in compressed sensing can be classified into two categories, iterative greedy pursuit and convex relaxation [26]. The way of greedy pursuit to recover the sparse signal is to identify the nonzero indices of the measured signal. The representative algorithm in this class is Orthogonal Matching

[1] Define ℓ_0-norm of the vector $\mathbf{s} \in R^N$ as the number of nonzero elements in $\mathbf{s}$.

Pursuit (OMP), which determines which column in the measurement matrix contribute to the measurement vector [27]. Specifically, given a m-sparsity signal $\mathbf{s}$, OMP aims to minimize the ℓ_2-norm [2] of the residue by selecting one column in each iteration, which is most correlated with the current residue, then being subtracted off to obtain the new residue for next iteration. After m iterations, the residue vanishes and the correct set of indices can be identified for the sparse signal recovery. On the other hand, convex relaxation techniques aims to relax the ℓ_0-norm problem to a ℓ_1-norm problem, whose solution is well known as Basis Pursuit (BP) [25]. Being very different from the standard least squares (LS) minimization procedures, Basis Pursuit is proposed as a convex alternative to the combinatorial norm ℓ_0 [29], which synthesizes a signal as a sparse superposition of waveforms by counting the nonzero entries of the vector [30]. However, given the measurement matrix $\mathbf{\Phi} \in \mathbb{C}^{N\times d}$ Gaussian independent and identical distributed (i.i.d) entries, BP requires at least $\mathcal{O}(N^2 d^{1.5})$ number of floating point operations [31], which prevents it from practical implementation due to the enormous computational consumption. At this point, BP is not suitable for large scale compressed sensing problem, i.e., massive MIMO scenario, whereas OMP attracts more interest for researcher due to its benchmark performance of BP and fast execution. Many variants of OMP have been proposed in recent years, e.g., order recursive least square matching pursuit (ORLSMP) [32], stagewise OMP (StOMP) [33], regularized OMP (ROMP) [34] and backtracking-based adaptive OMP (BAOMP) [35]. However, there are two drawbacks of the standard OMP methods: (1) it requires the knowledge of the sparsity m, and (2) it requires more measurements than BP. To overcome the limitation, Extended OMP is reported in [26] as two pieces of algorithms, OMP_α and OMP_∞. In [26],

[2] Define ℓ_p norm of the vector $\mathbf{s} \in R^N$ as $(\|\mathbf{s}\|)_p = \sum_{i=1}^{N} |s_i|^p$, and $\|s\|_2 = \|s\|$ for simplicity.

OMP_α method aims to reduce the number of measurements close to BP method, and to achieve superior performance as well by going beyond m-iterations. Meanwhile, the purpose of OMP_∞ is in order to get rid of the prior knowledge of the sparsity m by running OMP_α until the signal residue vanishes. Besides, generalized OMP (gOMP) is proposed to further improve the recovery performance and computational efficiency via choosing multiple indices in each iteration instead of the strongest component, but with the penalty of additional identification steps [36].

In addition, the spatial sparsity enables the possibility of deploying the hybrid MIMO instead of fully digital antenna arrays, and the hybrid beamforming and combining techniques have been a growing field of interest for researchers [16, 37–40]. Moreover, based on the sparsity in angle domain, the conventional channel estimation can be conducted as estimating the angle information (i.e. DoA and DoD) and the corresponding channel gain of each path. In [41], the authors applied the discrete Fourier transform (DFT) beamformers to construct the hybrid precoding structure and proposed a two-stage channel estimation strategy where the path directions were estimated via a two dimensional Multiple Signal Classification (MUSIC), which will be elaborated in the next section, and the path gains were estimated by the least squares method in the second stage. The DFT operation were also utilized in [16] to implement hybrid massive MIMO structure with limited RF chains, and further introduced an additional digital operation, namely spatial rotation, to concentrate all the energy on a single sample by shifting the projection of the channel response onto the Fourier basis, thus enormously reducing the demanded RF chains. Then a channel tracking method is proposed in [16] to show the near-optimal performance it can achieve with the least number of RF chains, where the angle information was tracked by a modified

unscented Kalman filter and the path gain was then estimated via a low-cost beam training towards the tracked DoA. This spatial rotation concept was further developed to implement non-orthogonal angle division multiple access (ADMA) for multiple users in [42], and to obtain better resolution of DoA estimation in [43] and [44].

Throughout this thesis, our focus is the hybrid massive MIMO structure with limited RF chains. As discussed in the previous chapter, channel tracking is demanding in hybrid massive MIMO, especially in high-mobility scenarios. The benefit coming with the systems with limited RF chains is the lower pilot overhead or even eliminating the pilot overhead by channel tracking or prediction. The purpose of channel tracking (or prediction) methods is to estimate the future CSI via the prior knowledge of current and past CSI fundamentally based on the principle of analyzing and forecasting the channel characteristics. To achieve this goal, there should be at least a set of parametric characteristics to be slowly varying in a specific domain, and can be represented in a model. The channel tracking procedure will be proposed based on such a model. To the best of our knowledge, the existing methods can be generally classified into: (a) Parametric Radio Channel model (PRC) (b) AutoRegressive (AR) model, (c) Basis-Expansion Model (BEM), which will be discussed in this section.

2.2.1 Parametric radio channel model (PRC)

Intuitively, if the channel model is deterministic by a set of parameters, it is natural to propose PRC methods to reconstruct the channel by estimating the deterministic parameters. The underlying principal of PRC methods is that the parameters vary much slower than the actual channel in a specific domain. Once these parameters are tracked, then

the channel response can be extrapolated for the future. As the mmWave massive MIMO communication evolves from the conventional Single-Input-Single-Output (SISO) point-to-point communication systems, the channel prediction has been exhaustively investigated in [45–48] based on the limited techniques about two decades ago. [46] and [47] model the fast fading SISO channel as a combination of multiple complex exponential functions with individual amplitudes, Doppler shifts and random phases, which is in terms of Jake's model. Then the channel response can be viewed as two parts: the amplitude and the complex exponential function. Thanks to the mature techniques in the array signal processing area, there are classical approaches to solve the frequency estimation problem, e.g. the Multiple Signal Classification (MUSIC) and the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [46, 49–51], which will be elaborated in the next section. In [46], the authors modified a variant of conventional MUSIC method, namely ROOT-MUSIC, to obtain the super-high resolution frequency estimation, and followed by a least squares fit to estimate the corresponding amplitude. In [47], the amplitudes are also estimated via a least squares solution, whereas the channel poles are jointly estimated by ESPRIT at different frequencies due to the wideband effect. However, channel prediction methods, especially for SISO channel are generally short-term tracking, and calibration is frequently demanded.

Before massive MIMO is formally proposed by Dr. Marzetta in 2010 [3], large-scale MIMO had already attracted great interest for researchers [52–55]. The channel prediction approaches for SISO were adapted to serve MIMO channel [56] and multi-user MIMO channel [57] by further utilizing the spatial structure of MIMO. As for mmWave massive MIMO communication, where the angle domain is much more sparse due to the severe

path loss, the channel can be described by the angle information and the complex gain. The angle information varies much slower than the channel gain, which implies the highly temporal and spatial channel correlation. In [58], the ULA (Uniform Linear Array) channel response is decomposed into angular information and beam gain. The angular information was obtained by aligning the beam towards the direction of signals, but suffering from the huge computational overhead, and then the beam gain was acquired by the linear Kalman Filter (KF) algorithm. [16] used similar methods but applied them in the opposite way, where a modified unscented KF was used to track the DoAs and a low-cost beam training estimated corresponding path gain. To greatly reduce the computational complexity of tracking methods, [59] was proposed to track the DoAs based on a physical movement model, following much smaller pilot training for path gain estimation. However, it turned out that the estimation accuracy it achieved can not support successfully uplink data detection or downlink data transmission. In comparison, [1] provides a sophisticated data-aided approach from system-level perspective. In [1], a data-aided channel tracking method was proposed to track the current channel based on the data detection results after channel decoding, where the angular information and the path gain were updated via a least squares solution, i.e. Newton's method. Since the channel tracking process highly replies on the data detection results, it is crucial to make data detection trustable. To this end, even with powerful channel coding theory, Inter-user Interference (IUI) is still a big issue to deal with. The problem is that [1] only investigated single-user scenario, and the closed-form expression of the CRLB was missing. As a result, we will extend the study to multi-user scenario in Chapter 4 in this thesis, and also give the closed-form CRLB of the single-user scenario.

2.2.2 Autoregressive model (AR)

AR model is another category of popular channel tracking schemes in the past two decades, which is based on the fact that the future behavior of the channel can be modeled as a weighted linear combination of the past channel estimates [60–64]. The underlying principle to determine the weights is to minimize the Mean Square Error (MSE), which requires the prior knowledge of the channel autocorrelation functions [65]. For SISO channels, to get the AR coefficients, the Burg method is applied in [62,66], whereas [63,67] investigated covariance and modified covariance approaches. However, even worse than the PRC methods, these simple schemes can hardly support short-term fast fading channel tracking due to the sensitivity to the noise and error propagation issue [67].

With the emergence of MIMO, temporal correlation was firstly investigated by assuming the MIMO channel as parallel SISO channels and utilizing individual MMSE predictors to track each SISO channel [68]. Then [69] further explored the spatial correlation between multiple sub-channels based on a vector AR model. However, this kind of multi-channel AR scheme suffers from the extremely high computational complexity, which hinders it from practical implementation. Inspired by the object localization and tracking solutions, the classical autoregressive methods, i.e. Kalman Filtering (KF) and Particle Filtering (PF), including their variants, have been a growing field of interest for researchers [70–75]. In [70–72], the space-time coding schemes were applied to model the MIMO channel as an autoregressive process and the KF was applied to track the channel information to enable space-time decoding successfully. But this kind of space-time coding schemes is only applicable for small-scale MIMO systems. As for large-scale MIMO

in [73], the authors propose a data-aided channel tracking algorithm for MIMO systems with VB (Variational Bayes) receivers, where a VB soft-input KF is applied to track the time-varying channel. However, the detected soft symbols may be unreliable to introduce outliers. This issue degrades the tracking performance, and requires additional training to improve the robustness.

However, AR channel tracking algorithms also have shortcomings, such as high estimation variance and error propagation. Since the underlying principle of AR model is to predict the future channel based on the knowledge of the past feature, error propagation problem is inevitable and the tracking results may not be reliable for the scenario of the fast fading channel with time-multiplexed training. The reason is that, in this scenario, we can only use the channel estimation results from the previous training session instead of the previous data detection session, thus leading to high estimation variance [76]. To solve this problem, various of data-aided approaches have been proposed to explore the KF by treating the data detection result as the measurement vector, and channel tracking and data detection can be jointly implemented by the extended KF [16, 77–81]. To further take spatial sparsity into account, [77] and [78] propose a standard KF-based channel tracking scheme to predict and modify the path gain of the LoS sub-channels according to the MMSE principle. By extensively exploring the spatial sparsity, many variants of the standard KF scheme are proposed to track partial channel information. For example, [79] applies an Extended Kalman Filtering (EKF) to track the DoA, whereas [16] and [80] both utilize a Unscented Kalman Filtering (UKF) to track the spatial angular information. Although the KF-based channel tracking method is well established in TDD (Time Division Duplex), [81] tackles this challenge in the FDD (Frequency Division Duplex) to further

reduce the training burden. As we can see, AR methods are commonly incorporated with PRC to output optimal or near-optimal performance by minimizing the MSE, and it is also recently combined with BEM methods, which will be discussed as following.

2.2.3 Basis Expansion Model (BEM)

As discussed previously, it is challenging to capture the features of fast-varying channels with PRC or AR models. To this end, BEM has been explored to describe the fast-varying channel taps as a superposition of series of time-varying basis functions to specifically model the Doppler effect, where the weights (or coefficients) are time-invariant [82]. To the best of our knowledge, there are several candidate basis functions, such as (Fourier) complex exponential (CE-BEM) [76, 80, 83], polynomial (P-BEM) [84, 85], wavelet (W-BEM) [86], and Slepian basis, i.e. Discrete Prolate Spheroidal Sequences (DPSS-BEM) [87, 88]. Different from the fast-varying channel taps, the coefficients of the BEM methods vary much slower that the channel taps, which makes itself possible to be tracked over each block time to predict the evolution of the fast-fading channel [76].

Among these, CE-BEM is the most popular one since the orthogonal Fourier basis is nature to describe the Doppler shift. In [83], a doubly-selective (i.e. frequency- and time-selective) fading SIMO (Single-Input-Multiple-Output) channel was modeled by CE-BEM with the coefficients tracked by KF, and a decision-directed channel tracking scheme was proposed using the estimated BEM coefficients. Since [89] has shown that the Fourier basis associated with rectangular windows suffered from the spectral leakage problem, another decision-directed channel tracking scheme was proposed in [84], based on a P-BEM decomposition, where the coefficients were tracked by Recursive Least Squares (RLS)

method within a sliding window for every block. Besides, to avoid the above frequency leakage problem, [86] and [87] exploited the orthonormal wavelets and DPSS respectively to represent the unknown channel time variations to compensate the multipath effects in CDMA (Code Division Multiple Access) systems, which is already outdated and out of our scope.

Regarding BEM-based channel prediction of MIMO channels, spatial and temporal correlation is still the most important property to enable the reduced-dimensional subspace channel tracking methods. [85] proposed a sparse channel tracking method for OFDM systems, where the path gain was tracked using P-BEM based on the fact that the angle and delay of each path changes slowly over a few adjacent OFDM symbols. In [80], the transmitted blocks were overlapped and the sub-block wise updating of the CE-BEM coefficients were perform through two scenarios: UKF for one-order AR model of basis coefficients, and RLS for basis coefficients without any assumed model. [88] also applied overlapped frames, but used DPSS-BEM to exploit the channel variation inside each sub-frame and an AR model was also proposed in [88] to tackled the scenario without considering any overlapping frames.

2.3 Methodologies

2.3.1 Receiver architectures

At the receiver side, the capacity can be achieved by joint Maximum Likelihood (ML) decoding of the data streams, but the complexity grows exponentially with the number of data streams. In this section, we introduce the fundamentals of linear receiver architectures,

which use linear operations to convert the problem of joint decoding all data streams into each separate decoding of individual data stream. These architectures extract the spatial degree of freedom gains, which has been introduced in the first chapter. In conjunction with an appropriate non-linear operation (successive cancellation of data streams), we can achieve the capacity of the fast fading massive MIMO channel.

Consider a MIMO channel model with N_t transmit antennas and N_r receive antennas, where $N_t \leq N_r$, and the transmitted data streams are independent. Then the received signal is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} = \sum_{k=1}^{N_t} \mathbf{h}_k x_k + \mathbf{w}, \tag{2.22}$$

where $\mathbf{h}_k$ denotes the k-th column of the channel matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, while x_k is the k-th symbol in data stream $\mathbf{x}$ and $\mathbf{w}$ is the CSCG noise. Apparently, different symbols are modulated on different columns of the channel matrix. Suppose that we can find a vector $\tilde{\mathbf{h}}_1$ such that

$$\tilde{\mathbf{h}}_1^H \mathbf{h}_1 = 1, \tag{2.23}$$

$$\tilde{\mathbf{h}}_1^H \mathbf{h}_k = 0, k = 2, 3, \cdots, N_t. \tag{2.24}$$

Therefore, x_1 can be estimated as

$$\hat{x}_1 = \tilde{\mathbf{h}}_1^H \mathbf{y} = \sum_{k=1}^{N_t} \tilde{\mathbf{h}}_1^H \mathbf{h}_k x_k + \tilde{\mathbf{h}}_1^H \mathbf{w} = x_1 + \tilde{\mathbf{h}}_1^H \mathbf{w}. \tag{2.25}$$

As we can see, by projecting the received signal on $\tilde{\mathbf{h}}_1$, the interference from other antennas

can be eliminated. Similarly, we can find $\tilde{\mathbf{h}}_k$ for $k = 2, \cdots, N_t$ such that

$$\tilde{\mathbf{h}}_k^H \mathbf{h}_1 = 1, \tag{2.26}$$

$$\tilde{\mathbf{h}}_k^H \mathbf{h}_{k'} = 0, k \neq k'. \tag{2.27}$$

Then x_k can be detected as

$$\hat{x}_k = \tilde{\mathbf{h}}_k^H \mathbf{y} = \sum_{k'=1}^{N_t} \tilde{\mathbf{h}}_k^H \mathbf{h}_{k'} x_{k'} + \tilde{\mathbf{h}}_k^H \mathbf{w} = x_k + \tilde{\mathbf{h}}_k^H \mathbf{w}. \tag{2.28}$$

It is promising that the inter-symbol interference is totally eliminated. If we construct a matrix $\mathbf{H}^\dagger = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_{N_t}\right]^H$, we have $\mathbf{H}^\dagger \mathbf{H} = \mathbf{I}_{N_t}$. After that, the transmitted data sequence can be detected as

$$\hat{\mathbf{x}} = \tilde{\mathbf{H}} \mathbf{y} = \mathbf{x} + \tilde{\mathbf{H}} \mathbf{w}. \tag{2.29}$$

This kind of detector is referred to as the *linear decorrelator* or *zero-forcing (ZF)* detector since it can completely eliminate the interference among different data streams. To this end, we can obtain the linear decorrelation matrix as

$$\mathbf{H}^\dagger = \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H. \tag{2.30}$$

This is based on the assumption of perfect CSI, which is not available in practice. Besides, it has poor performance in low SNR regime. The noise vector is $\tilde{\mathbf{w}} = \mathbf{H}^\dagger \mathbf{w}$, and its covariance matrix is given by

$$E\left\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\right\} = \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H E\left\{\mathbf{w}\mathbf{w}^H\right\} \mathbf{H} \left(\mathbf{H}^H \mathbf{H}\right)^{-1} = N_0 \left(\mathbf{H}^H \mathbf{H}\right)^{-1}, \tag{2.31}$$

where N_0 is the variance of the CSCG noise.

Apart from ZF detector, the Matched Filter (MF) is also commonly used to maximize the output SNR. In this case, for the k-th user, we first project the received signal on $\mathbf{h}_k^H$ as

$$\tilde{x}_k = \mathbf{h}_k^H \mathbf{y} = \|\mathbf{h}_k\|^2 x_k + \sum_{k' \neq k} \mathbf{h}^H \mathbf{h}_{k'} x_{k'} + \mathbf{h}_k^H \mathbf{w}. \tag{2.32}$$

Then x_k can be estimated by normalizing $\tilde{x}_k$ by a factor of $\|\mathbf{h}_k\|^2$ as

$$\hat{x}_k = \frac{\tilde{\mathbf{x}}_k}{\|\mathbf{h}_k\|^2} = x_k + \sum_{k' \neq k} \frac{\mathbf{h}_k^H \mathbf{h}_{k'}}{\|\mathbf{h}_k\|^2} x_{k'} + \frac{\mathbf{h}_k^H \mathbf{w}}{\|\mathbf{h}_k\|^2}. \tag{2.33}$$

In the matrix form, we have the estimate of $\mathbf{x}$ as

$$\hat{\mathbf{x}} = \left(\text{diag}\{\mathbf{H}^H \mathbf{H}\}\right)^{-1} \mathbf{H}^H \mathbf{y}. \tag{2.34}$$

As we can see, the linear decorrelator aims to eliminate interference, while the matched filter targets to suppress noise. However, an optimal detector should be able to jointly minimize the total influence of the noise and interference. To be specific, we should find a matrix $\mathbf{A}$ to estimate data sequence $\mathbf{x}$ as

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{y}, \tag{2.35}$$

so that

$$\mathbf{a}_k = \arg\min_{\mathbf{a}_k} E_{\mathbf{w}} \left\{ \|\mathbf{a}_k^H \mathbf{y} - \mathbf{x}\|^2 \right\}. \tag{2.36}$$

The MMSE (Minimum Mean Square Error) receiver of the k-th user is given by

$$\mathbf{a}_k = \left(\sum_{k' \neq k} P_{k'} \mathbf{h}_{k'} \mathbf{h}_{k'}^H + N_0 \mathbf{I}_{N_t} \right)^{-1} \mathbf{h}_k. \tag{2.37}$$

The comparison among matched filter, linear decorrelator and MMSE detector is shown in Figure 2.3. Theoretically, MMSE receive always outperforms the other two linear receivers. In low SNR regime, the MF receiver outperforms the decorrelator, and is comparable to the MMSE receiver. The reason is that the interference is negligible in this case, and noise dominates. On the other hand, in high SNR regime, the decorrelator outperforms the MF receiver, and is comparable to the MMSE receiver. The reason is that noise negligible in this case and interference dominates. As we can see in Figure 2.3, it seems to be different from the ideal case because we only simulate one single user in the cell. In our simulation case, QPSK (Quatrature Phase Shift Keying) is conducted as modulation method, and there is one user with single antenna and one BS with single antenna as well. Single tap channel is assumed and SNR range is $[-10, 20]$ dB. All the solid lines indicate the simulation results, and all dots indicate the theoretical results, which are well matched. The yellow color represents the theoretical and simulation results with AWGN (Additive White Guassian Noise) channel, which is obviously the best performance we can approach while dealing with the fading channel. The green and blue ones represent the BER (Bit Error Rate) performance with MF and ZF detectors respectively, which are also overlapped with each other. The reason is that there is only one user without inter-user interference where ZF is equivalent to MF detection. Last but not the least, the red one indicates that MMSE receiver always outperforms the others.

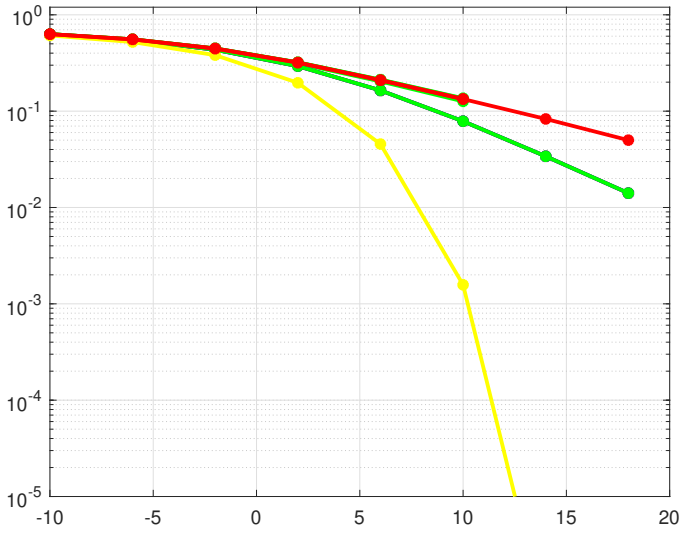

Figure 2.3: Comparison of BER performance.

2.3.2 DoA estimation

Throughout this thesis, our focus is only on the first category of the channel tracking classification, PRC model. Therefore, some mature DoA estimation methods will be introduced in this section, such as MUSIC (Multiple Signal Classification) and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques). Aforementioned codebook scheme for beam alignment and compressive sensing family achieve the similar goal. Given the spatial information, complex gain can be obtained with much smaller pilot training, which dramatically reduces the overall system overhead. We should note here that the Doppler effect estimation is excluded in this thesis by assuming the Doppler effect is perfectly eliminated via PLLs at the RF end, which is not our interest. Besides, this thesis focus on the LoS path scenario due to the characteristics of mmWave communication. As

for multipath scenario, there are some subspace dimension estimation methods, such as AIC (Akaike Information Criterion) and MDL (Minimum Description Length), to estimate the number of path in multiple resolvable path scenario, which is out of scope of this thesis [90, 91].

To introduce MUSIC algorithm, firstly let us assume a point-to-point MIMO system with ULA antenna arrays consisting of K and M antenna elements at the transmitter and receiver respectively. Then the received signal is given as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{2.38}$$

where $\mathbf{H} \in \mathbb{C}^{M\times K}$ denotes the complex channel matrix, and $\mathbf{z}$ is the CSCG noise with variance σ^2 following $\mathbb{N}(0, \sigma^2\mathbf{I})$. Here we assume $\mathbf{H}$ is with full column rank of K, and each column is given as $\mathbf{h}_k = \beta_k \mathbf{a}(\theta_k)$, where β_k denotes the complex gain and θ_k is the DoA of the signal from the k-th transmit antenna. The autocorrelation of the far-field received signal is given by

$$\mathbf{R}_{yy} = E\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^H + \mathbf{R}_{zz}, \tag{2.39}$$

where $\mathbf{R}_{xx}$ represents the autocorrelation of the transmitting signal and $R_{zz} = \sigma^2\mathbf{I}$ denotes the noise autocorrelation. At this point, Eigenvalue Decomposition can be conducted as

$$\mathbf{R}_{yy} = \sum_{i=1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^H, \tag{2.40}$$

where λ_i are eigenvalues of the autocorrelation of the received signal such that $\lambda_1 \geq \lambda_2 \geq$

$\cdots \geq \lambda_K \geq \sigma^2$. At this end, (2.40) can be rewritten as

$$\mathbf{R}_{yy} = \begin{bmatrix} \mathbf{E}_x & \mathbf{E}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_s & \\ & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{E}_s^H \\ \mathbf{E}_n^H \end{bmatrix} = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^H + \mathbf{E}_n \mathbf{\Lambda}_n \mathbf{E}_n^H, \tag{2.41}$$

where $\mathbf{E}_s$ and $\mathbf{E}_n$ denote eigenvectors of the signal space and noise space respectively, the diagonal elements in $\mathbf{\Lambda}_s = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_K \end{bmatrix}$ and $\mathbf{\Lambda}_n = \sigma^2 \mathbf{I}_t$ where $\mathbf{I}_t \in \mathcal{R}^{(M-K)\times(M-K)}$ denote the eigenvalues of the signal space and noise space respectively. The key idea of the MUSIC algorithm is that the signal subspace and the noise subspace are orthogonal to each other

$$\mathbf{E}_n^H \mathbf{a}(\theta_i) = 0, i = 1, \ldots, K \tag{2.42}$$

Therefore, the orthogonality of the signal subspace and noise subspace results in the fact that the array steering vector $\mathbf{a}(\theta_i)$ in the direction of a source is orthogonal to the noise eigenvectors as shown in Equation (2.42). This results in peaks in the plot of the MUSIC pseudo-spectrum, which is given by

$$P_{MUSIC}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta)}. \tag{2.43}$$

The DoA θ estimates are obtained either by manually locating the peaks of Equation (2.43) or using a search algorithm to identify the peaks

$$\theta_i = \arg\min_{\theta} \mathbf{a}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\mathbf{a}(\theta). \tag{2.44}$$

To sum up, the process of the DoA estimation based on MUSIC algorithm is as follows

- Suppose Q observations have been received, and the estimated covariance matrix is given by
$$\hat{\mathbf{R}}_{yy} = \frac{1}{Q}\sum_{q=1}^{Q}\mathbf{y}_q\mathbf{y}_q^H, \tag{2.45}$$
- Conduct EVD on the estimated covariance matrix $\hat{\mathbf{R}}_{yy} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$,
- According to the descending order of the eigenvalues in $\boldsymbol{\Sigma}$, find the last $M - K$ eigenvalues as the noise subspace eigenvalues and the corresponding eigenvectors in $\mathbf{U}$ to construct $\mathbf{E}_n$ which is given by Equation (2.40).
- Based on Equation (2.43), use a searching algorithm with respect to θ to find the peaks of the MUSIC pseudo-spectrum to identify the DoA estimate.

The performance of the MUSIC algorithm and its suitability for practical applications is limited by the brute-force searching dimension. Since the spacing of θ to locate the peak is critical to the estimation accuracy, the searching step is required to be as fine as possible. However, this will cause extremely high computational complexity which prevents MUSIC for the practical applications. To overcome this limitation, Root-MUSIC was proposed as a modified version of MUSIC based on polynomial root finding. Root-MUSIC has been shown to increase the resolution and decrease the computational complexity of MUSIC algorithm. Root-MUSIC is also a model based algorithm which directly estimates the DoA parameter by explicitly using a model of the array steering vector and the received data.

The ESPRIT algorithm is another mature subspace method to estimate DoA and possibly the most popular in channel prediction studies. Compared with MUSIC, it offers more

robustness against array imperfections and since it does not require brute-force search, it has significantly lower computational complexity and storage requirement. The main idea of ESPRIT is the shift invariance structure of the array obtained by creating two overlapping sub-arrays from the original array with a constant translation distance.

Suppose the same scenario as aforementioned MUSIC example, the Vandermonde structured array steering matrix $\mathbf{A} \in \mathbb{C}^{M \times K}$ is defined as

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-2)\omega_1} & e^{j(M-2)\omega_2} & \dots & e^{j(M-2)\omega_K} \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_2} & \dots & e^{j(M-1)\omega_K} \end{bmatrix}. \tag{2.46}$$

where the steering vector $\mathbf{e}[\omega]$ is conventionally defined as $\mathbf{e}[\omega_i] = \left[1, e^{j\omega_i}, ..., e^{j(M-1)\omega_i}\right]^T$ with $\omega_= \frac{2\pi d \cos\theta}{\lambda}$. Here d is the distance between any adjacent antenna elements and θ denotes the Direction of Arrival (DoA) of the incoming signal. After that, based on the shift invariance structure, $\mathbf{A}$ can be split into two submatrices $\mathbf{A}_1 \in \mathbb{C}^{(M-1)\times K}$ and $\mathbf{A}_2 \in \mathbb{C}^{(M-1)\times K}$ as

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-2)\omega_1} & e^{j(M-2)\omega_2} & \dots & e^{j(M-2)\omega_K} \end{bmatrix}, \tag{2.47}$$

$$\mathbf{A}_2 = \begin{bmatrix} e^{j\omega_1} & e^{j\omega_2} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(M-2)\omega_1} & e^{j(M-2)\omega_2} & \dots & e^{j(M-2)\omega_K} \\ e^{j(M-1)\omega_1} & e^{j(M-1)\omega_2} & \dots & e^{j(M-1)\omega_K} \end{bmatrix}. \tag{2.48}$$

Apparently, we can see that there is a relationship between two submatrices $\mathbf{A}_2 = \mathbf{A}_1\mathbf{\Phi}$, where $\mathbf{\Phi}$ is termed the rotational operator as defined as

$$\mathbf{\Phi} = \text{diag}\left\{e^{j\omega_1}, e^{j\omega_2}, \cdots, e^{j\omega_K}\right\}. \tag{2.49}$$

Intuitively, all the DoA information can be determined once we know the rotational operator $\mathbf{\Phi}$. Therefore, the goal of ESPRIT method is to determine $\mathbf{\Phi}$ by explicitly applying the spatial array response and its shift invariance structure. Firstly, we split the received signal $\mathbf{Y} = \beta\mathbf{A}\mathbf{x} + \mathbf{n}$ as

$$\mathbf{Y}_1 = \beta\mathbf{A}_1\mathbf{x}_1 + \mathbf{n}_1, \tag{2.50}$$

$$\mathbf{Y}_2 = \beta\mathbf{A}_2\mathbf{x}_2 + \mathbf{n}_2, \tag{2.51}$$

where β is the complex gain, $\mathbf{x}_1 = [x_1, x_2, \cdots, x_{K-1}]$ and $\mathbf{x}_2 = [x_2, x_3, \cdots, x_K]$ are the subsequences of the original transmitted data sequence, and $\mathbf{n}_1 \in \mathbb{C}_{(M-1)\times 1}$ and $\mathbf{n}_2 \in \mathbb{C}_{(M-1)\times 1}$ denote the CSCG noise vectors. Since we only focus on DoA estimation, we simplify the notation by getting rid of the complex gain β which can be estimated by a small pilot overhead once we estimate DoA accurately. At this point, we can construct a new structure of

the received signal $\mathbf{Z} \in \mathbb{C}^{2(M-1)\times 1}$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{x}_1 + \mathbf{n}_1 \\ \mathbf{A}_2\mathbf{x}_2 + \mathbf{n}_2 \end{bmatrix} = \tilde{\mathbf{A}}\mathbf{x} + \mathbf{N}, \tag{2.52}$$

where $\tilde{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_1\mathbf{\Phi} \end{bmatrix}$ and $\mathbf{N} = \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{bmatrix}$. Similar to MUSIC method, based on the concept of subspace, the autocorrelation of the new structure of the received signal is given by

$$\mathbf{R}_{ZZ} = E\left[\mathbf{Z}\mathbf{Z}^H\right] = \tilde{\mathbf{A}}\mathbf{R}_{xx}\tilde{\mathbf{A}}^H + \sigma^2\mathbf{I}, \tag{2.53}$$

where R_{xx} represents the autocorrelation of the transmitting signal and σ denotes the variance of the noise. Then the EVD of the autocorrelation can be also written as the similar form as Equation (2.41), with respect to the signal subspace and noise subspace

$$\mathbf{R}_{ZZ} = \sum_{i=1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^H = \mathbf{E}_s\mathbf{\Lambda}_s\mathbf{E}_s^H + \mathbf{E}_n\mathbf{\Lambda}_n\mathbf{E}_n^H. \tag{2.54}$$

There must exist a unique nonsingular matrix $\mathbf{T}$ satisfying

$$\mathbf{E}_s = \tilde{\mathbf{A}}\mathbf{T} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{T} \\ \mathbf{A}_1\mathbf{\Phi}\mathbf{T} \end{bmatrix}. \tag{2.55}$$

Intuitively, $\mathbf{E}_1$ is related to $\mathbf{E}_2$ from Equation (2.55)

$$\mathbf{E}_2 = \mathbf{E}_1\mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T} = \mathbf{E}_1\mathbf{\Psi}, \tag{2.56}$$

where $\mathbf{\Psi} = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}$. Apparently, diagonal elements in $\mathbf{\Phi}$ are the eigenvalues of the matrix $\mathbf{\Psi}$. That is equivalent to say, once we obtain $\mathbf{\Psi}$, we can estimate all the DoA information via EVD of $\mathbf{\Psi}$ to further obtain $\mathbf{\Phi}$.

To sum up, the process of the DoA estimation based on ESPRIT algorithm is as follows

- For each observation $\mathbf{Y}$, construct the new received signal $\mathbf{Z}$ based on Equation (2.52)
- Suppose Q observations have been received, and the estimated covariance matrix is given by

$$\hat{\mathbf{R}}_{ZZ} = \frac{1}{Q}\sum_{q=1}^{Q}\mathbf{Z}_q\mathbf{Z}_q^H, \tag{2.57}$$

- Conduct EVD of $\hat{\mathbf{R}}_{ZZ}$ to obtain the signal subspace matrix $\mathbf{E}_s$,
- Apply Least Square algorithm to obtain estimated $\hat{\mathbf{\Psi}}$ based on Equation (2.56),
- Conduct EVD of $\hat{\mathbf{\Psi}}$ to obtain $\hat{\mathbf{\Phi}}$ to identify DoA information which are the eigenvalues of $\hat{\mathbf{\Psi}}$ or equivalently the diagonal elements of $\hat{\mathbf{\Phi}}$.

2.4 Summary

In this chapter, we firstly introduced the fundamentals of wireless channel characteristics, massive MIMO and millimeter wave communications. Specifically, kinds of fading due to multi-path and Doppler effects are elaborated, degree of freedom in massive MIMO and spatial sparsity in mmWave communication are also carefully discussed. After that, some existing work in acquiring the CSI in mmWave massive MIMO communication systems are reviewed, where we showed the limitations and deficiencies of the conventional channel estimation and tracking methods. In the remaining chapters of this thesis, we will present

the adaptive data-aided channel tracking methods for hybrid mmWave massiveMIMO systems with different antenna types, and multi-user scenario is also taken into account for DLA case.

Chapter 3

ULA-based data-aided channel tracking framework

As the first main chapter of this thesis, in order to explore different aspects of a data-aided channel tracking system, we will start with the simple case which is based on Uniform Linear Array (ULA) for single user scenario. In this chapter, to analyze the different influences of various parameters on channel tracking accuracy, it is necessary to ignore the inter-user interference by only considering single user scenario. In the next chapter, multi-user case will be explored.

3.1 ULA-based channel response

As for ULA-based communication system, the angle space is equally spaced by $2\pi/M$ where M is the number of antennas. Here we assume massive MIMO is applied to compensate the huge pass loss of mmWave communication, where M is generally large (e.g. at least

hundreds of antennas). In addition, TDD is employed to exploit the channel reciprocity, and only uplink channel estimation will be conducted. To be specific, as shown in Figure 3.1, the BS is equipped with ULA with M antennas and there is only one user in the cell with single antenna. Assume that the ULA antenna elements are equally spaced by distance $d = 2\pi/M$, the indices is defined as $I_m \in \{0, ..., M-1\}$, and the 0-th antenna is set to be the reference point, as indicated in Figure 3.1. Assume the Direction of Arrival (DoA) θ of the uplink received signal is defined with respect to the left hand side, and the channel vector of the user with slow fading is given as

$$\mathbf{h} = \beta e^{j\phi} \mathbf{e}[\omega], \tag{3.1}$$

where β is the attenuation factor which is inversely proportional to the distance between the user and the BS, ϕ denotes the random phase delay and the steering vector $\mathbf{e}[\omega]$ is conventionally defined as $\mathbf{e}[\omega] = \left[1, e^{j\omega}, ..., e^{j(M-1)\omega}\right]^T$ with $\omega = \frac{2\pi d \cos\theta}{\lambda}$. Therefore, the channel response on the m-th antenna element is given by

$$\mathbf{h}[m] = \beta e^{j(m\omega+\phi)}. \tag{3.2}$$

It should be noted here that β is a real number, which differs from the equivalent complex gain which will be introduced in the second part of this chapter with DLA array.

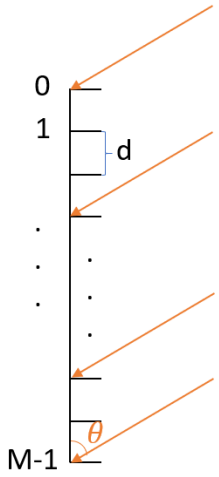

Figure 3.1: ULA architecture.

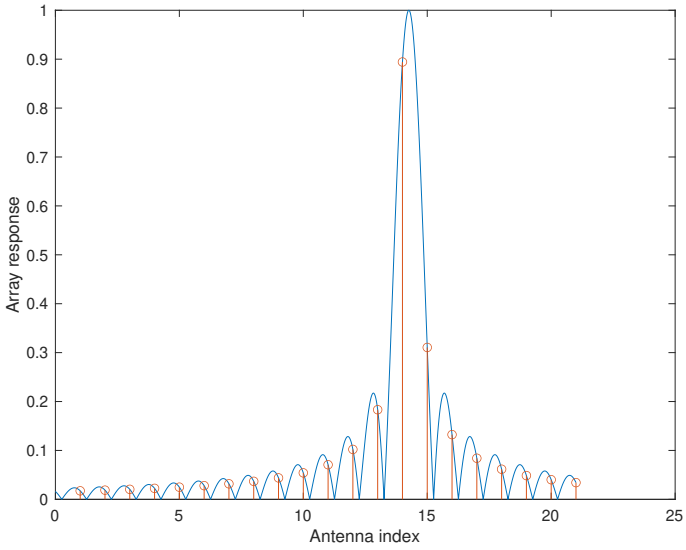

Figure 3.2: spatial sparsity.

3.2 Data-aided channel tracking system model

To evaluate the performance, we follow the conventional data-aided channel tracking protocol, which is illustrated in Figure 3.3 [1]. At the beginning of the 0-th time slot, pilot-based channel estimation is required by sending the orthogonal pilot sequence $\mathbf{p}$ via uplink channel

$$\mathbf{Y}_0 = \mathbf{h}_0\mathbf{p}^T + \mathbf{N}_0, \tag{3.3}$$

where $\mathbf{h}_0$ denotes the channel vector at the 0-th time slot and each element of $\mathbf{N}_0$ is a CSCG variable with variance σ^2. Then in every following time slots, there are typically three phases to conduct, namely processing, downlink data transmission and uplink data detection. In the first processing phase at the 1-st time slot, initial channel estimation is conducted

$$\hat{\mathbf{h}}_0 = \mathbf{Y}_0\mathbf{p}^*/\|\mathbf{p}\|^2 = \mathbf{h}_0 + \mathbf{W}_0, \tag{3.4}$$

where $\mathbf{W}_0 = \mathbf{N}_0/\|\mathbf{p}\|^2$ is still a CSCG matrix with variance σ^2. It is clear that the estimate is only contaminated by the noise by conducting the orthogonal pilot sequence. Then this initial estimate will be used for downlink data beamforming and uplink data detection within this 1-st time slot. By the end of the 1-st time slot, the decoded data stream will be used for updating the channel estimation in the next processing phase. This step is called data-aided channel tracking. As shown in Figure 3.3, as long as this process iterates successfully, the overhead of the conventional channel estimation via sending pilots for every time slot is eliminated. To be specific, both downlink and uplink data processing at any $(n+1)$-th time slot utilize the channel estimate from the previous time slot. Figure 3.4

shows a detailed block diagram for a conventional data-aided channel tracking system [1]. At any $(n+1)$-th time slot, the received uplink signal $\mathbf{Y}_{n+1} \in \mathbb{C}^{M\times N}$ is

$$\mathbf{Y}_{n+1} = \mathbf{h}_{n+1}\mathbf{x}_{n+1}^{T} + \mathbf{N}_{n+1}, \tag{3.5}$$

where $\mathbf{h}_{n+1} \in \mathbb{C}^{M\times 1}$ denotes the channel vector at the $(n+1)$-th time slot, $\mathbf{x}_{n+1} \in \mathbb{C}^{N\times 1}$ is the transmitted data vector with length of N, and $\mathbf{N}_{n+1}$ is additive Gaussian noise, in which each element is i.i.d. CSCG random variables with variance σ^2.

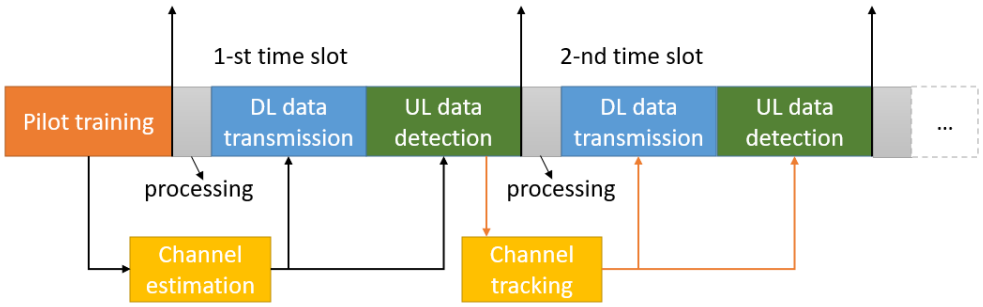

Figure 3.3: The data flow of the data-aided channel tracking system.

Based on the fact that the channel is highly correlated between adjacent time slots, i.e. $\mathbf{h}_{n+1} \approx \mathbf{h}_n$, the uplink data stream can be detected by the matched filter based on the channel estimate from the previous time slot

$$\tilde{\mathbf{x}}_{n+1} = \mathbf{Y}_{n+1}^{T}\hat{\mathbf{h}}_n^{*}/\|\hat{\mathbf{h}}_n\|^2. \tag{3.6}$$

After the powerful channel decoder, an accurate estimate of the data stream $\hat{\mathbf{x}}_{n+1}$ can

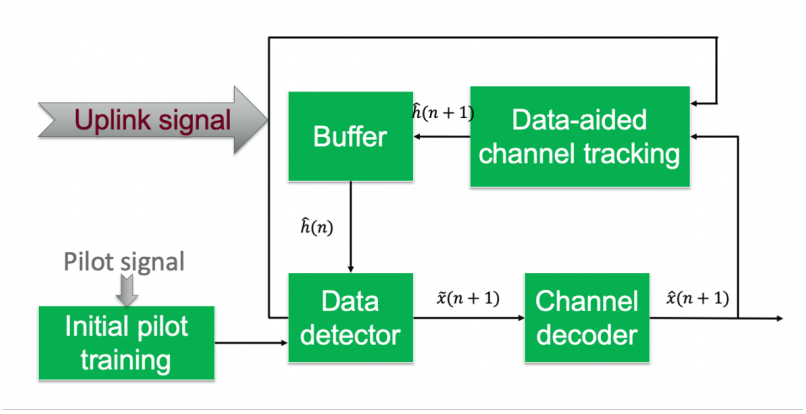

Figure 3.4: The system diagram of the data-aided channel tracking system [1].

be obtained without most of the errors, which will be further used to perform data-aided channel tracking

$$\hat{\mathbf{h}}_{n+1} = \mathbf{Y}_{n+1}\hat{\mathbf{x}}^*_{n+1}/\|\hat{\mathbf{x}}_{n+1}\|^2. \tag{3.7}$$

This tracking result will be updated as the current estimate $\hat{\mathbf{h}}_n$ to be used for the next time slot. As long as this tracking process iterates fast enough, we will reduce significant overhead coming from the pilot-based channel estimation for every time slot.

3.3 CRLB analysis

3.3.1 CRLB analysis with spatial signature estimation

In this section, we will discuss the CRLB of the estimation error to evaluate the performance. Due to the fact of spatial sparsity, to estimate the $\hat{\mathbf{h}}$ is equivalent to estimate ω, ϕ

and β in this case. The probability density function (PDF) of channel estimate $\hat{\mathbf{h}}$ is

$$f(\hat{\mathbf{h}}|\omega,\phi,\beta) = (\pi\sigma^2)^{-M}\exp\left\{-\frac{\|\hat{\mathbf{h}}-\mathbf{h}\|^2}{\sigma_e^2}\right\}, \tag{3.8}$$

where σ_e^2 is the variance of estimation error, following zero-mean Gaussian distribution as

$$E\{(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\} = \sigma_e^2\mathbf{I}_M. \tag{3.9}$$

Then the log-likelihood function will be given as

$$l(\hat{\mathbf{h}}|\omega,\phi,\beta) = \ln f = -M\ln(\pi\sigma_e^2) - \frac{\|\hat{\mathbf{h}}-\mathbf{h}\|^2}{\sigma_e^2}. \tag{3.10}$$

The partial derivatives of l with respect to the estimated parameters are given as

$$\begin{aligned}
\frac{\partial l}{\partial \omega} &= \frac{2}{\sigma_e^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial \mathbf{h}_r}{\partial \omega} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial \mathbf{h}_i}{\partial \omega}\right],\\
\frac{\partial l}{\partial \phi} &= \frac{2}{\sigma_e^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial \mathbf{h}_r}{\partial \phi} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial \mathbf{h}_i}{\partial \phi}\right],\\
\frac{\partial l}{\partial \beta} &= \frac{2}{\sigma_e^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial \mathbf{h}_r}{\partial \beta} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial \mathbf{h}_i}{\partial \beta}\right],,
\end{aligned} \tag{3.11}$$

where $\mathbf{h}$ is a combination of the real part and the imaginary part as $\mathbf{h} = \mathbf{h}_r + \mathbf{h}_i$. In this case, the Fisher information matrix is simply given as

$$\mathbf{F} = E\left\{\begin{bmatrix} \left(\frac{\partial l}{\partial \omega}\right)^2 & \frac{\partial l}{\partial \omega}\frac{\partial l}{\partial \phi} & \frac{\partial l}{\partial \omega}\frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \omega} & \left(\frac{\partial l}{\partial \phi}\right)^2 & \frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \omega} & \frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \phi} & \left(\frac{\partial l}{\partial \beta}\right)^2 \end{bmatrix}\right\}, \tag{3.12}$$

and the expectations of these quadratic terms are given by

$$
\begin{aligned}
E\left\{\left(\frac{\partial l}{\partial \omega}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \omega}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \omega}\right|^2\right),\\
E\left\{\left(\frac{\partial l}{\partial \phi}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \phi}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \phi}\right|^2\right),\\
E\left\{\left(\frac{\partial l}{\partial \beta}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \beta}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \beta}\right|^2\right),\\
E\left\{\frac{\partial l}{\partial \omega}\frac{\partial l}{\partial \phi}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \omega}\frac{\partial \mathbf{h}_r}{\partial \phi} + \frac{\partial \mathbf{h}_i^T}{\partial \omega}\frac{\partial \mathbf{h}_i}{\partial \phi}\right),\\
E\left\{\frac{\partial l}{\partial \omega}\frac{\partial l}{\partial \beta}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \omega}\frac{\partial \mathbf{h}_r}{\partial \beta} + \frac{\partial \mathbf{h}_i^T}{\partial \omega}\frac{\partial \mathbf{h}_i}{\partial \beta}\right),\\
E\left\{\frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \beta}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \phi}\frac{\partial \mathbf{h}_r}{\partial \beta} + \frac{\partial \mathbf{h}_i^T}{\partial \phi}\frac{\partial \mathbf{h}_i}{\partial \beta}\right).
\end{aligned}
\tag{3.13}
$$

Here, let $\mathbf{\Phi} = [\omega, \phi, \beta]^T$. The Fisher information matrix can be rewritten as

$$
\mathbf{F} = \frac{2}{\sigma^2}\mathbf{G}^T\mathbf{G}, \tag{3.14}
$$

where the Jacobi matrix $\mathbf{G}$ is given as

$$
\mathbf{G} = \left[\frac{\partial \mathbf{h}_r^T}{\partial \mathbf{\Phi}}, \frac{\partial \mathbf{h}_i^T}{\partial \mathbf{\Phi}}\right]^T. \tag{3.15}
$$

Moreover, for the m-th antenna element, the partial derivatives are given as

$$
\begin{aligned}
\frac{\partial \mathbf{h}_r[m]}{\partial \omega} &= -m\beta\sin(m\omega+\phi),\ \frac{\partial \mathbf{h}_i[m]}{\partial \omega} = m\beta\cos(m\omega+\phi),\\
\frac{\partial \mathbf{h}_r[m]}{\partial \phi} &= -\beta\sin(m\omega+\phi),\ \frac{\partial \mathbf{h}_i[m]}{\partial \phi} = \beta\cos(m\omega+\phi),\\
\frac{\partial \mathbf{h}_r[m]}{\partial \beta} &= \cos(m\omega+\phi),\ \frac{\partial \mathbf{h}_i[m]}{\partial \beta} = \sin(m\omega+\phi).
\end{aligned}
\tag{3.16}
$$

By substituting the above partial derivatives (3.16) into equation (3.12), the Fisher matrix becomes

$$\mathbf{F} = \frac{2}{\sigma_e^2}\begin{bmatrix} \beta^2 \sum_{m=0}^{M-1} m^2 & \beta^2 \sum_{m=0}^{M-1} m & 0 \\ \beta^2 \sum_{m=0}^{M-1} m & \beta^2 M & 0 \\ 0 & 0 & M \end{bmatrix}, \tag{3.17}$$

To derive the inverse of the Fisher matrix, we denote $\mathbf{A} = \beta^2 \begin{bmatrix} \sum_{m=0}^{M-1} m^2 & \sum_{m=0}^{M-1} m \\ \sum_{m=0}^{M-1} m & M \end{bmatrix}$, $\mathbf{B} = \mathbf{C}^T = [0,0]^T$ and $\mathbf{D} = [M]$ to facilitate Block Matrix Inversion theorem, which is given as

$$\mathbf{F}^{-1} = \frac{2}{\sigma_e^2}\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \frac{2}{\sigma_e^2}\begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}. \tag{3.18}$$

Then we can get the CRLB with respect to ω

$$\begin{aligned} \mathbf{F}^{-1}[1,1] &= \frac{\sigma^2}{2\beta^2}\left(\sum_{m=0}^{M-1} m^2 - \sum_{m=0}^{M-1} m\frac{1}{M}\sum_{m=0}^{M-1} m\right)^{-1} \\ &= \frac{\sigma^2}{2\beta^2}\left(\frac{(M-1)M(2M-1)}{6} - \frac{M(M-1)^2}{4}\right)^{-1} \\ &= \frac{\sigma^2}{2\beta^2}\left(\frac{M^3+5M}{12}\right)^{-1}. \end{aligned} \tag{3.19}$$

The CRLB with respect to ϕ can be derived as

$$\begin{aligned} \mathbf{F}^{-1}[2,2] &= \frac{\sigma^2}{2\beta^2}\left(\frac{1}{M} - \frac{1}{M}\sum_{m=0}^{M-1} m\left(\frac{M^3+5M}{12}\right)^{-1}\sum_{m=0}^{M-1} m\frac{1}{M}\right) \\ &= \frac{\sigma^2}{2\beta^2}\left(\frac{2M^2}{M^3+6M^2+3M}\right). \end{aligned} \tag{3.20}$$

Last, the CRLB with respect to β will be

$$\mathbf{F}^{-1}[3,3] = \frac{\sigma^2}{2\beta^2 M}. \tag{3.21}$$

Based on the fact that M is sufficiently large, we can conclude

$$|\Delta\omega|^2 \geq \mathbf{F}^{-1}[1,1] \sim \frac{6\sigma^2}{\beta^2 M^3}, \tag{3.22}$$

$$|\Delta\phi|^2 \geq \mathbf{F}^{-1}[2,2] \sim \frac{\sigma^2}{\beta^2 M}, \tag{3.23}$$

$$|\Delta\beta|^2 \geq \mathbf{F}^{-1}[3,3] = \frac{\sigma^2}{2\beta^2 M}. \tag{3.24}$$

We can see that the lower bounds of the tracking errors associated with ω, ϕ and β are all proportional to the variance of the noise σ^2, which is equivalent to say that it is inversely proportional to the length of the data sequence N. Therefore, it can improve the channel tracking accuracy by increasing the data block size. From equations (3.22), (3.23) and (3.24), we can also see that all lower bounds of the parameters leading channel tracking error are inversely proportional to the attenuation factor β, which is intuitive due to the fact that the stronger the LoS path is, the better data detection performance that is achieved, thus leading to better channel tracking accuracy. Last but not the least, we can conclude that ω, ϕ and β are all inversely proportional to the total number of the antennas M, which means larger antenna arrays help reduce the estimation error, especially with respect to the estimation error of the spatial signature ω.

3.3.2 CRLB analysis with DoA estimation

Furthermore, based on the relationship between the spatial signature ω and the DoA θ, it seems to be equivalent to estimate the DoA directly. Next, we will show the CRLB with respect to the θ itself. In this case, we assume the base station is still equipped with massive ULA, where M antennas are critically spaced by $d = \lambda/2$. The channel vector of the single user under slow fading in this scenario is given as

$$\mathbf{h} = \beta e^{j\phi}\mathbf{e}[\theta], \tag{3.25}$$

where the steering vector is with respect to θ directly as $\mathbf{e} = \left[1, e^{j\pi\cos\theta}, ..., e^{j(M-1)\pi\cos\theta}\right]^T$. Therefore, the channel response on the m-th antenna element is given by

$$\mathbf{h}[m] = \beta e^{j(m\pi\cos\theta+\phi)}. \tag{3.26}$$

By applying the aforementioned channel tracking method to get $\hat{\mathbf{h}}$, the probability density function $f\left(\hat{\mathbf{h}}|\theta\right)$ and the corresponding likelihood function $l\left(\hat{\mathbf{h}}|\theta\right)$ are very similar to equations (3.8) and (3.10) with different parameters

$$f(\hat{\mathbf{h}}|\theta,\phi,\beta) = \left(\pi\sigma^2\right)^{-M}\exp\left\{-\frac{\|\hat{\mathbf{h}}-\mathbf{h}\|^2}{\sigma_e^2}\right\}, \tag{3.27}$$

$$l(\hat{\mathbf{h}}|\theta,\phi,\beta) = \ln f = -M\ln\left(\pi\sigma_e^2\right) - \frac{\|\hat{\mathbf{h}}-\mathbf{h}\|^2}{\sigma_e^2}. \tag{3.28}$$

Following a similar derivation as in the previous section, the Fisher information matrix will be very similar to (3.12) only by substituting $\frac{\partial l}{\partial \omega}$ with $\frac{\partial l}{\partial \theta}$

$$\mathbf{F} = E\left\{\begin{bmatrix} \left(\frac{\partial l}{\partial \theta}\right)^2 & \frac{\partial l}{\partial \theta}\frac{\partial l}{\partial \phi} & \frac{\partial l}{\partial \theta}\frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \theta} & \left(\frac{\partial l}{\partial \phi}\right)^2 & \frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \theta} & \frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \phi} & \left(\frac{\partial l}{\partial \beta}\right)^2 \end{bmatrix}\right\}, \tag{3.29}$$

The partial derivative of l with respect to θ is given as

$$\frac{\partial l}{\partial \theta} = \frac{2}{\sigma^2}\left[(\hat{\mathbf{h}}_r - \mathbf{h}_r)^T \frac{\partial \mathbf{h}_r}{\partial \theta} + (\hat{\mathbf{h}}_i - \mathbf{h}_i)^T \frac{\partial \mathbf{h}_i}{\partial \theta}\right], \tag{3.30}$$

where $\mathbf{h} = \mathbf{h}_r + \mathbf{h}_i$. At this point, for the m-th antenna element, the partial derivative becomes

$$\begin{aligned} &\frac{\partial \mathbf{h}_r[m]}{\partial \theta} = \beta m\pi \sin\theta \sin\left(m\pi\cos\theta + \phi\right), \ \frac{\partial \mathbf{h}_i[m]}{\partial \theta} = -\beta m\pi \sin\theta \cos\left(m\pi\cos\theta + \phi\right), \\ &\frac{\partial \mathbf{h}_r[m]}{\partial \phi} = -\beta \sin\left(m\pi\cos\theta + \phi\right), \ \frac{\partial \mathbf{h}_i[m]}{\partial \phi} = \beta \cos\left(m\pi\cos\theta + \phi\right), \\ &\frac{\partial \mathbf{h}_r[m]}{\partial \beta} = \cos\left(m\pi\cos\theta + \phi\right), \ \frac{\partial \mathbf{h}_i[m]}{\partial \beta} = \sin\left(m\pi\cos\theta + \phi\right). \end{aligned} \tag{3.31}$$

Similar to the previous case, the Fisher information matrix is simply given as

$$\mathbf{F} = \frac{2}{\sigma_e^2}\begin{bmatrix} \beta^2\pi^2 \sin^2\theta \sum_{m=0}^{M-1} m^2 & \beta^2\pi \sin\theta \sum_{m=0}^{M-1} m & 0 \\ \beta^2\pi \sin\theta \sum_{m=0}^{M-1} m & \beta^2 M & 0 \\ 0 & 0 & M \end{bmatrix}, \tag{3.32}$$

Based on the Block Matrix Inversion theorem in (3.18), the inversion of the Fisher matrix

can be derived. Therefore, the CRLB with respect to the DoA θ is

$$\begin{aligned}\mathbf{F}^{-1}[1,1] &= \frac{\sigma^2 M}{2\beta^2}\left(\pi^2\sin^2\theta M\sum_{m=0}^{M-1}m^2-\pi^2\sin^2\theta\left(\sum_{m=0}^{M-1}m\right)^2\right)^{-1}\\ &= \frac{\sigma^2}{2\beta^2}\left(\frac{(M-1)M(2M-1)}{6}-\frac{1}{M}\left(\frac{M(M-1)}{2}\right)^2\right)^{-1}\\ &= \frac{\sigma^2}{6\beta^2\pi^2\sin^2\theta(M^3-M)}.\end{aligned} \tag{3.33}$$

Then the CRLB with respect to ϕ is derived as

$$\begin{aligned}\mathbf{F}^{-1}[2,2] &= \frac{\sigma^2\pi^2\sin^2\theta}{2\beta^2}\left(M\sum_{m=0}^{M-1}m^2-\left(\sum_{m=0}^{M-1}m\right)^2\right)^{-1}\pi^2\sin^2\theta\sum_{m=0}^{M-1}m^2\\ &= \frac{\sigma^2}{2\beta^2}\left(M-\frac{\left(\sum_{m=0}^{M-1}m\right)^2}{\sum_{m=0}^{M-1}m^2}\right)^{-1}\\ &= \frac{\sigma^2(2M-1)}{\beta^2(-M^2+2M)}.\end{aligned} \tag{3.34}$$

After that, the CRLB with respect to β is derived as

$$\mathbf{F}^{-1}[3,3] = \frac{\sigma^2}{2\beta^2 M}. \tag{3.35}$$

Since M is generally sufficiently large in mMIMO communication, it is concluded that

$$|\Delta\theta|^2 \geq \mathbf{F}^{-1}[1,1] \sim \frac{\sigma^2}{6M^3\beta^2\pi^2\sin^2\theta}, \tag{3.36}$$

$$|\Delta\phi|^2 \geq \mathbf{F}^{-1}[2,2] \sim \frac{\sigma^2}{\beta^2 M}, \tag{3.37}$$

$$|\Delta\beta|^2 \geq \mathbf{F}^{-1}[3,3] = \frac{\sigma^2}{2\beta^2 M}. \tag{3.38}$$

As we can see from equations above (3.37) and (3.38), the impact on the estimation error associated with ϕ and β is exactly the same with those in the previous case when we estimate the spatial signature instead of DoA, which are illustrated by equations (3.23) and (3.24). They are both inversely proportional to the length of data sequence N, attenuation factor β and the number of antennas M of the ULA. This also means that we can improve the estimation accuracy by increasing the data block size and / or increase massive MIMO scale. However, different from estimating the spatial signature as a whole, the lower bound directly with respect to the DoA depends on the function '$\sin\theta$' in equation (3.36), which is a nonlinear function. Intuitively, the estimation accuracy will be better when the incoming signal arrives at the position around the center beam while the estimation accuracy will be worse when it falls into the side beams. Since the estimation accuracy is not stable, tracking the spatial signature is a better choice rather than the DoA itself by applying traditional array signal processing techniques.

3.4 Simulation results and outcome analysis

For the proposed data aided channel tracking system with ULA, there are some important parameters, the data block length N, the number of antennas at the BS/RSU (Road Side Unit) and SNR. Based on the above theoretical analysis, in this section, we will conduct simulations to evaluate the impacts of these parameters on channel tracking performance with spatial signature estimation.

To set up the whole system, we assume TDD (Time Division Duplex) to explore channel reciprocity, and single user scenario. Therefore, we only need to conduct uplink channel

estimation, and downlink data transmission is able to apply the same estimate via feedback channel with negligible delay. Be consistent with the aforementioned setup environment, the BS/RSU is deployed with large ULA with M antenna elements, and only one user with single antenna is moving within the cell with speed of 360 km per hour, or equivalently 100 meter per second, which is the typical speed for the Chinese high-speed railway system. Throughout the simulation, assume the modulation scheme is defaulted to be QPSK, the block length of data transmission is N and the symbol duration is 1 ms. The length of the training pilot is set to be 100 to get decent initial channel estimation, and the path loss exponent is chosen to be 2 due to the LoS path only. First of all, Figure 3.5 shows

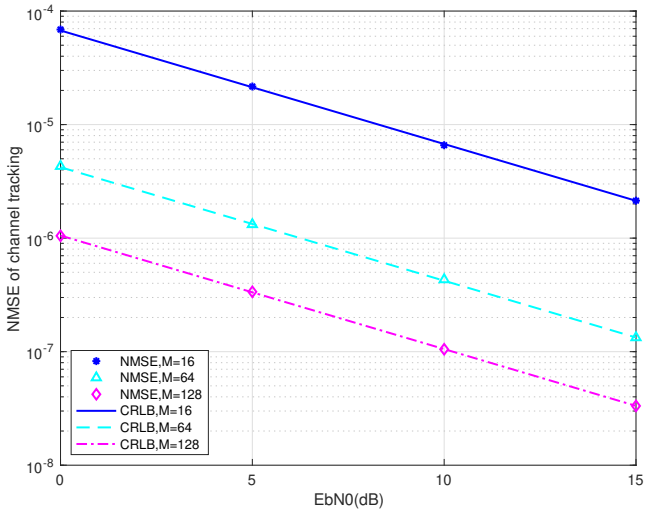

Figure 3.5: ULA-based channel tracking error with spatial signature estimation.

the preliminary results of the system performance at the 100-th time slot, in terms of the NMSE (Normalized Mean Square Error) of the channel tracking estimates. The SNR range is $[0, 15]$ dB and the number of antenna elements varies as $[16, 64, 128]$. As we can

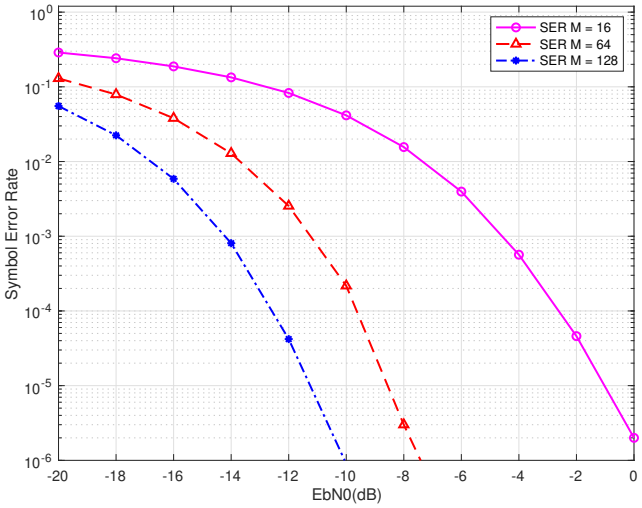

Figure 3.6: Power gain for different numbers of antenna elements.

see from Figure 3.5, the simulation results match the corresponding CRLBs very well. The reason is that the CRLB is derived without the data detection error, even though both data detection error and noise contribute to the channel estimation error. However, throughout the thesis, the detected data sequence is assumed to be trustable via channel coding scheme. In our simulation, multiple antennas at the receiver side provide some degrees of freedom, all of which are used to enhance the diversity gain due to the single user scenario. Figure 3.6 illustrates the performance enhancement on SER with the power gain of M. By applying MF detector, the achieved SNR is boosted by a factor of M (i.e., the receiving antenna number), thus improving the SER performance of data detection. In Figure 3.6, the SER is still satisfactory for down to -20 dB. Therefore, for SNR range of $[0, 15]$ dB, the data detection error is negligible in our simulation, which is the reason why numerical results match CRLBs such well. We can conclude that the tracking error is

dominated by the noise and the data detection error contributes little to channel tracking inaccuracy in this case.

Next we will demonstrate the other impacts on the accuracy performance associated with parameters M, in terms of the number of antennas. As we can see from Figures 3.5 and 3.7, where the numbers of antennas are varying from $16 * [1, 4, 8]$ and $64 * [1, 4, 8]$ respectively, the NMSE trends are nearly identical. As we increase M four times, the accuracy can be enhanced by at least one order of magnitude. The close-form CRLBs of ULA-based channel tracking errors in equations (3.22) to (3.24) are all inversely proportional to M, which is verified by our simulation results. This is quite intuitive that more antennas at receiver side provide more power gain to improve the data detection accuracy, thus boosting the channel tracking performance. However, the spatial beams will be narrower due to a large antenna array, which leads to more challenging to estimate of DoA.

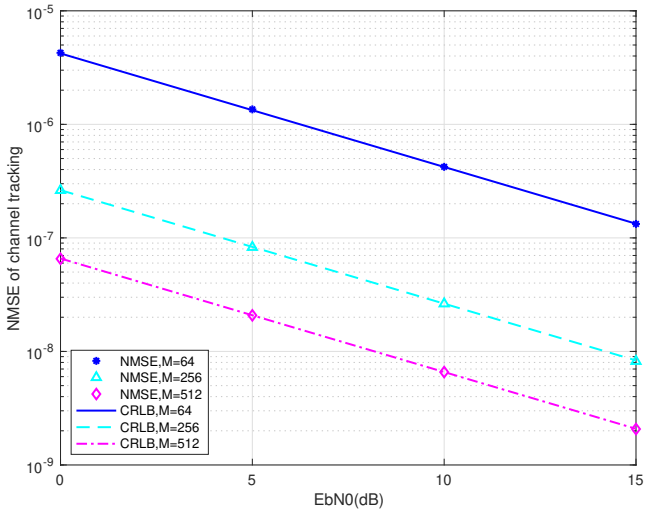

Figure 3.7: Impact of antenna number M on channel tracking error.

Last but not the least, we will see the impact of the block length N on the channel tracking performance in Figure 3.8. Also recall from the close-form CRLBs in equations (3.22) to (3.24) that CRLBs are proportional to the noise variance σ^2, while the longer data sequence helps to suppress the noise. Every time we double the block length N, the NMSE increases by a factor of two. On the other hand, we are not able to use a very large block length N since we need to use the previous channel estimates to approximate the current CSI to facilitate data-aided channel tracking process. As a result, the channel states of adjacent time slots have to be highly correlated, which requires a reasonable block length.

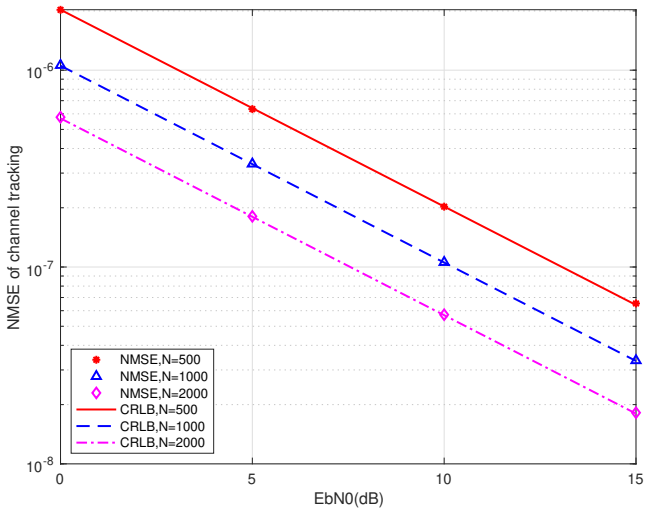

Figure 3.8: Impact of block length N on channel tracking error.

To sum up, based on the aforementioned theoretical analysis and simulation outcomes, our closed-form CRLB expressions have been verified. The tracking error decreases with

more antenna elements. Moreover, the accuracy performance is also boosted with larger block lengths. However, the block length also needs to be carefully chosen based on the time correlation assumption. The last point is that estimating spatial signature is always better than tracking the DoA for channel tracking systems.

3.5 Summary

In this chapter, a ULA-based data-aided channel tracking framework is presented for mmWave massive MIMO system. The critical assumption is that users are well separated in space and Doppler effect is complementary at the RF end. The conventional data-aided channel tracking method is proposed with Newton iteration algorithm. To evaluate the accuracy performance, thorough CRLB analysis is achieved by deriving the closed-form expression of CRLBs. This CRLB analysis also proves that spatial signature estimation we applied is the better choice than the traditional DoA estimation for channel prediction solutions. From the simulation results, we can see that the proposed data-aided channel tracking accuracy closely approaches the CRLB. To reduce the cost and complexity, we will investigate DLA-based system model with limited RF chains.

Chapter 4

DLA-based data-aided channel tracking framework

In this chapter, the channel tracking method for DLA-based mmWave massive MIMO system is adopted from [1] where the close form expression of the CRLB is missing. In this thesis, the closed-form expression of the CRLB is derived and the multi-user scenario is further explored in the next chapter.

4.1 DLA-based channel response

Generally, a DLA consists of an electromagnetic lens and an antenna array whose elements are located on the focal arc of the lens in the azimuth plane. As shown in Figure 4.1, the EM lens is of size $D_y \times D_z$ with negligible thickness, the radius F is defined as the focal length, $\theta_m \in [-\pi/2, \pi/2]$ is the angle of the m-th antenna element relative to the x-axis, within the range of $-(M-1)/2 \leq m \leq (M-1)/2$ when M is odd, ϕ_p is the physical arrival

angle of one specific signal path, which is denoted as the angle between incident wave and x-axis, and the normalized dimension of the lens D is predefined as $D \triangleq D_y/\lambda$, where λ is the carrier wavelength. Among these, the number of antennas on the focal arc M is related with D via $M = 1 + 2\lfloor D \rfloor$. Without loss of generality, M antennas are critically spaced on the focal arc, or equivalently $\tilde{\theta}_m \triangleq \sin\theta_m$ are equally spaced in the interval $[-1, 1]$. Different from ULAs, the antenna elements are placed much denser around the center of the DLA than those at the two ends of the focal arc as shown in Figure 4.1. For the remaining of this thesis, we call the beams at the center and at the two ends of the focal arc as *central beams* and *side beams* respectively, and call the antenna elements around the 0-th antenna element and around the $\pm\left(\frac{M-1}{2}\right)$-th antenna element as *central antennas* and *side antennas* as well.

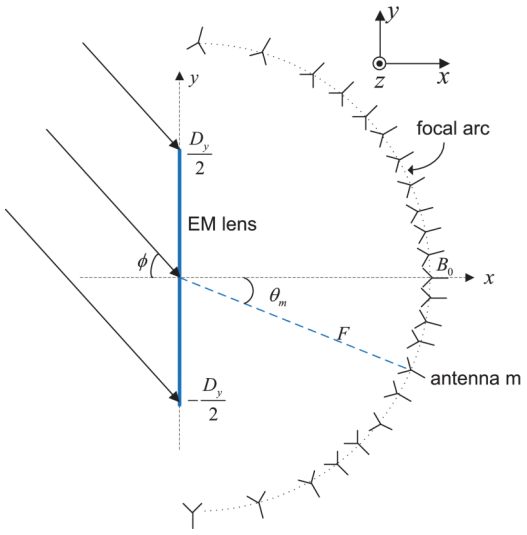

Figure 4.1: DLA architecture [2].

Based on the optical characteristics of the EM lens, the waveform of the incoming

signal is collected by the lens, and then concentrated onto a single point of the continuous focal arc. However, as only limited antenna elements are placed on the focal arc, DLA can collect discrete samples of the received waveform. The authors in [2] proved that the array response of a DLA approximates the DFT of the array response of a Uniform Linear Array (ULA), thus following 'sinc' function in angle domain by exploring the spatial sparsity.

Theorem 2. *[2] For a DLA with critically spaced antenna elements, assuming the DoA of the incoming signal is ϕ, the array response $\mathbf{a}(\phi)$ follows 'sinc' function via approximating DFT operation on the array response of a ULA, and can be expressed as*

$$\mathbf{a}(\phi) = \text{sinc}\left(m - D(1+\cos\phi)\right), \ \forall m \in [0, M-1]. \tag{4.1}$$

Proof: Provided in Appendix A.

4.2 Truncated channel vector

Considering the similar property in spatial spectrum, it can be easily seen that the received signal will be distributed onto the entire antenna array following the 'sinc' waveform. This fact means that the energy on the two antenna elements in the main lobe is strong while the others is much weaker. In [24], it is shown that over 80% energy concentrates on the two samples in the main lobe, which enables the truncated channel vector with limited RF chains. Instead of activating all the antenna elements with enormous energy consumption, only a small number of antennas are required to be powered on to serve each user. Assume that V antenna elements are associated to be active to collect signal from a single user,

then denote $\mathbf{h}_{t,k}$ as the truncated channel vector of the k-th user

$$\mathbf{h}_{t,k} = \mathbf{h}_k[v_k : v_k + V - 1], \tag{4.2}$$

which means V consecutive antenna indices are chosen around the peak sample of the original channel vector. Algorithm 1 shows the method to find consecutive indices based on the peak index with the strongest energy. It has been proved that the energy loss ratio is less than 20% with only two active antennas for each user [1]. This shows that it is greatly promising to employ DLA with limited RF chains for V2I mmWave communication systems to dramatically reduce the cost and complexity.

Algorithm 1 Single-user channel truncating algorithm.

1: Find the peak index $m_{p,k}$ by the entire channel vector $\mathbf{h}_k$ by

$$m_{p,k} = \arg\max_{m} |\mathbf{h}_k[m]|; \tag{4.3}$$

2: Find the starting index v_k
3: **if** V is odd **then**
4: let $l = (V-1)/2$
5: **if** $m_{p,k} - l \leq \min(m),\ \forall m \in [0, M-1]$ **then**
6: $v_k = \min(m)$
7: **else if** $m_{p,k} + l \geq \max(m),\ \forall m \in [0, M-1]$ **then**
8: $v_k = \max(m) - 2l$
9: **else**
10: $v_k = m_{p,k} - l$
11: **end if**
12: **else if** V is even **then**
13: let $l = \begin{cases} V/2 - 1, & m_{p,k} \leq \alpha_k \\ V/2, & m_{p,k} > \alpha_k \end{cases}$
14: Repeat commands 5 – 11
15: **end if**
16: **return** $\mathbf{I}_k = [v_k : v_k + V - 1]$ and $\mathbf{h}_{k,t} = \mathbf{h}_k[\mathbf{I}_k]$;

4.3 System model

Assume in a mmWave massive MIMO system, DLA is equipped at the BS with M antenna elements via limited RF chains, which means only V antenna elements are activated for each user at each time slot, and the number of single-antenna users is one for single user scenario. Similar with [1], the channel vector is given as

$$\mathbf{h} = \beta \mathbf{a}(\phi), \tag{4.4}$$

where β is the complex gain, ϕ is the DoA of the incoming signal, and $\mathbf{a}(\phi)$ is the array response with the m-th element given by

$$\mathbf{a}(\phi)[m] = \text{sinc}(m - \alpha), \tag{4.5}$$

where $\alpha = D(1 + \cos\phi)$ and D is the normalized lens dimension satisfying $M = 1 + 2\lfloor D \rfloor$. By adapting truncated channel vector concept with V number of active antenna elements, substituting the truncated channel vector into aforementioned equations (3.3) to (3.7), and applying Least-square iterative channel tracking method in [1], we will get the estimated truncated channel vector $\hat{\mathbf{h}}_d$ with three estimation parameters α, β_r and β_i, where $\beta = \beta_r + \beta_i$. It should be noted that though the procedure is quite similar to the previous one with ULA, truncated channel vector concept is adopted here to save cost and complexity. After that, we will follow the similar analysis in [1] to derive the closed-form CRLB of the tracking error. However, in [1], the authors did not give the closed-form expression of the CRLB. We will give the detailed derivation of closed-form CRLB expression, which is one

of the major contribution of this thesis.

4.4 CRLB analysis

4.4.1 CRLB analysis with spatial signature estimation

Firstly, we rewrite $\hat{\mathbf{h}}_d$ as $\hat{\mathbf{g}}$ to get rid of the subscript, where $\hat{\mathbf{g}} = \hat{\mathbf{g}}_r + \hat{\mathbf{g}}_i$ and $\mathbf{g} = \mathbf{g}_r + \mathbf{g}_i$. The probability density function of $\hat{\mathbf{g}}$ is given as

$$f(\hat{\mathbf{g}}|\alpha,\beta) = (\pi\sigma_e^2)^{-V} \exp\left\{-\frac{\|\hat{\mathbf{g}}-\mathbf{g}\|^2}{\sigma_e^2}\right\}, \tag{4.6}$$

where σ_e^2 is the variance of estimation error, following zero-mean Gaussian distribution as

$$E\{(\hat{\mathbf{g}}-\mathbf{g})(\hat{\mathbf{g}}-\mathbf{g})^H\} = \sigma_e^2\mathbf{I}_V. \tag{4.7}$$

Then the log-likelihood function will be given as

$$l(\hat{\mathbf{g}}|\alpha,\beta) = \ln f = -V\ln\left(\pi\sigma_e^2\right) - \frac{\|\hat{\mathbf{g}}-\mathbf{g}\|^2}{\sigma_e^2} = -V\ln\left(\pi\sigma_e^2\right) - \frac{\|\hat{\mathbf{g}}_r-\mathbf{g}_r\|^2}{\sigma_e^2} - \frac{\|\hat{\mathbf{g}}_i-\mathbf{g}_i\|^2}{\sigma_e^2}. \tag{4.8}$$

According to the Fisher information matrix

$$\mathbf{F} = E\left\{\begin{bmatrix} \left(\frac{\partial l}{\partial\alpha}\right)^2 & \frac{\partial l}{\partial\alpha}\frac{\partial l}{\partial\beta_r} & \frac{\partial l}{\partial\alpha}\frac{\partial l}{\partial\beta_i} \\ \frac{\partial l}{\partial\beta_r}\frac{\partial l}{\partial\alpha} & \left(\frac{\partial l}{\partial\beta_r}\right)^2 & \frac{\partial l}{\partial\beta_r}\frac{\partial l}{\partial\beta_i} \\ \frac{\partial l}{\partial\beta_i}\frac{\partial l}{\partial\alpha} & \frac{\partial l}{\partial\beta_i}\frac{\partial l}{\partial\beta_r} & \left(\frac{\partial l}{\partial\beta_i}\right)^2 \end{bmatrix}\right\}, \tag{4.9}$$

we derive the first-order partial derivatives as following

$$\begin{aligned}
\frac{\partial l}{\partial \alpha} &= \frac{2}{\sigma_e^2}\left[(\hat{\mathbf{g}}_r - \mathbf{g}_r)^T \frac{\partial \mathbf{g}_r}{\partial \alpha} + (\hat{\mathbf{g}}_i - \mathbf{g}_i)^T \frac{\partial \mathbf{g}_i}{\partial \alpha}\right], \\
\frac{\partial l}{\partial \beta_r} &= \frac{2}{\sigma_e^2}(\hat{\mathbf{g}}_r - \mathbf{g}_r)^T \frac{\partial \mathbf{g}_r}{\partial \beta_r}, \\
\frac{\partial l}{\partial \beta_i} &= \frac{2}{\sigma_e^2}(\hat{\mathbf{g}}_i - \mathbf{g}_i)^T \frac{\partial \mathbf{g}_i}{\partial \beta_i},
\end{aligned} \tag{4.10}$$

with the partial derivative of each antenna element with respect to the estimation parameters given by

$$\begin{aligned}
\frac{\partial \mathbf{g}_r[v]}{\partial \alpha} &= \beta_r \left[\frac{-\pi \cos\left(\pi(v-\alpha)\right)}{\pi(v-\alpha)} + \frac{\pi \sin\left(\pi(v-\alpha)\right)}{\left(\pi(v-\alpha)\right)^2}\right] \\
&= \beta_r \frac{\mathrm{sinc}(v-\alpha) - \cos\left(\pi(v-\alpha)\right)}{v-\alpha}, \\
\frac{\partial \mathbf{g}_i[v]}{\partial \alpha} &= \beta_i \frac{\mathrm{sinc}(v-\alpha) - \cos\left(\pi(v-\alpha)\right)}{v-\alpha}, \\
\frac{\partial \mathbf{g}_r[v]}{\partial \beta_r} &= \frac{\partial \mathbf{g}_i[v]}{\partial \beta_i} \\
&= \mathrm{sinc}(v-\alpha), \\
\frac{\partial \mathbf{g}_r[v]}{\partial \beta_i} &= \frac{\partial \mathbf{g}_i[v]}{\partial \beta_r} \\
&= 0,
\end{aligned} \tag{4.11}$$

where $v \in \mathbf{I}_V$, $\mathbf{I}_V$ is the index set of active antennas. Furthermore we also derive the

expectations of the corresponding quadratic terms as shown

$$
\begin{aligned}
E\left\{\left(\frac{\partial l}{\partial \alpha}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{g}_r}{\partial \alpha}\right|^2 + \left|\frac{\partial \mathbf{g}_i}{\partial \alpha}\right|^2\right), \\
E\left\{\left(\frac{\partial l}{\partial \beta_r}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{g}_r}{\partial \beta_r}\right|^2\right), \\
E\left\{\left(\frac{\partial l}{\partial \beta_i}\right)^2\right\} &= \frac{2}{\sigma_e^2}\left(\left|\frac{\partial \mathbf{g}_i}{\partial \beta_i}\right|^2\right), \\
E\left\{\frac{\partial l}{\partial \alpha}\frac{\partial l}{\partial \beta_r}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{g}_r^T}{\partial \alpha}\frac{\partial \mathbf{g}_r}{\partial \beta_r}\right), \\
E\left\{\frac{\partial l}{\partial \alpha}\frac{\partial l}{\partial \beta_i}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{g}_i^T}{\partial \alpha}\frac{\partial \mathbf{g}_i}{\partial \beta_i}\right), \\
E\left\{\frac{\partial l}{\partial \beta_r}\frac{\partial l}{\partial \beta_i}\right\} &= \frac{2}{\sigma_e^2}\left(\frac{\partial \mathbf{g}_r^T}{\partial \beta_r}\frac{\partial \mathbf{g}_r}{\partial \beta_i}\right).
\end{aligned}
\tag{4.12}
$$

Here, let $\mathbf{\Psi} = [\alpha, \beta_r, \beta_i]^T$, the Fisher information matrix can be rewritten as

$$
\mathbf{F} = \frac{2}{\sigma_e^2}\mathbf{G}^T\mathbf{G}, \tag{4.13}
$$

where the Jacobi matrix $\mathbf{G}$ is given as

$$
\mathbf{G} = \left[\frac{\partial \mathbf{g}_r^T}{\partial \mathbf{\Psi}}, \frac{\partial \mathbf{g}_i^T}{\partial \mathbf{\Psi}}\right]^T. \tag{4.14}
$$

To be specific, the rewritten Fisher information matrix is given as

$$
\mathbf{F} = \frac{2}{\sigma_e^2}\begin{bmatrix} \|\beta\|^2\Phi & \beta_r\Omega & \beta_i\Omega \\ \beta_r\Omega & \Sigma & 0 \\ \beta_i\Omega & 0 & \Sigma \end{bmatrix}, \tag{4.15}
$$

where $\Phi = \sum_{v=v_0}^{v_0+V-1} \left(\frac{\mathrm{sinc}(v-\alpha) - \cos(\pi(v-\alpha))}{v-\alpha} \right)^2$, $\Omega = \sum_{v=v_0}^{v_0+V-1} \frac{\mathrm{sinc}^2(v-\alpha) - \mathrm{sinc}(v-\alpha)\cos(\pi(v-\alpha))}{v-\alpha}$, and $\Sigma = \sum_{v=v_0}^{v_0+V-1} \mathrm{sinc}^2(v-\alpha)$. To explore the CRLB of the estimation error of each parameter, we need to derive the inverse of Fisher information matrix based on the block matrix inversion theorem, which is given as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}, \quad (4.16)$$

Here we substitute $\mathbf{A} = \begin{bmatrix} \|\beta\|^2 \Phi & \beta_r \Omega \\ \beta_r \Omega & \Sigma \end{bmatrix}$, $\mathbf{B} = [\beta_i \Omega, 0]^T$, $\mathbf{C} = \mathbf{B}^T$, and $\mathbf{D} = \Sigma$. Then we can derive the diagonal elements of the inverse of the Fisher information matrix are given as

$$\mathbf{F}^{-1}[1,1] = \frac{\sigma^2}{2\|\beta\|^2 \left(\Phi - \frac{\Omega^2}{\Sigma}\right)}, \quad (4.17)$$

$$\mathbf{F}^{-1}[2,2] = \frac{\sigma^2 \left(\|\beta\|^2 \Phi\Sigma - \beta_i^2 \Omega^2\right)}{2\Sigma \left(\|\beta\|^2 \Phi\Sigma - \|\beta\|^2 \Omega^2\right)}, \quad (4.18)$$

$$\mathbf{F}^{-1}[3,3] = \frac{\sigma^2 \left(\|\beta\|^2 \Phi\Sigma - \beta_r^2 \Omega^2\right)}{\Sigma \left(\|\beta\|^2 \Phi\Sigma - \|\beta\|^2 \Omega^2\right)}. \quad (4.19)$$

Unfortunately, because of these complex summations Σ, Φ and Ω, it is hard to see the relations between tracking error and associated parameters. Therefore, the difficult task to deriving the closed-form CRLB of estimation errors on DLA is to derive the closed-form expressions of complex summations.

It is shown in Appendix B, based on two facts that the number of antenna elements is relatively large and the truncated channel vector preserves more than 80% of total energy,

we can approximate Φ, Ω and Σ based on above DFT properties as follows

$$\Phi \approx \frac{\pi^2 \left(M^3\right)}{3}, \tag{4.20}$$

$$\Omega \approx 0, \tag{4.21}$$

$$\Sigma \approx M. \tag{4.22}$$

Thus we can rewrite Equation (4.15) as

$$\mathbf{F} = \frac{2}{\sigma_e^2} \begin{bmatrix} \|\beta\|^2 \frac{\pi^2 M^3}{3} & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & M \end{bmatrix}, \tag{4.23}$$

Based on this beautiful Fisher information matrix, the CRLBs of estimation errors of each parameter, i.e., equations (4.17) to (4.19), will be rewritten as

$$|\Delta\alpha|^2 \geq \mathbf{F}^{-1}[1,1] \sim \frac{3\sigma_e^2}{2\pi^2 M^3 \|\beta\|^2}, \tag{4.24}$$

$$|\Delta\beta_r|^2 \geq \mathbf{F}^{-1}[2,2] \sim \frac{\sigma_e^2}{2M}, \tag{4.25}$$

$$|\Delta\beta_i|^2 \geq \mathbf{F}^{-1}[3,3] \sim \frac{\sigma_e^2}{2M}. \tag{4.26}$$

Compared to (3.22) to (3.24) in the previous section with ULA, we can see the relationships are very similar. Clearly, the lower bounds of the parameters leading channel tracking errors are all proportional to the variance of the noise σ^2, which is equivalent to say that they are inversely proportional to the length of the data sequence N. Therefore, it can improve the channel tracking accuracy by increasing the data block size. Secondly, the lower bound

of the parameter α is inversely proportional to the attenuation factor β, which is intuitive due to the fact that the stronger the LoS path is, the better data detection performance is achieved, thus leading to better channel tracking accuracy. Last but not the least, we can conclude that α and β are both inversely proportional to the total number of the antennas M, which means larger antenna arrays help reduce the estimation error, especially with respect to the estimation error of the spatial signature α. Equivalently, for the communication systems with limited RF chains, larger V active antennas for each user should help to reduce the estimation error.

4.4.2 CRLB analysis with DoA estimation

The second half of the section is similar to the counterpart of ULA. We will show the CRLB with respect to ϕ itself instead of estimating α. In this case, the aforementioned process is the same but only substituting α with ϕ. However, the corresponding partial derivatives

of each antenna element will be definitely different

$$\begin{aligned}
\frac{\partial \mathbf{g}_r[v]}{\partial \phi} &= \beta_r D \sin\phi \frac{\operatorname{sinc}\left(v - D(1+\cos\phi)\right) - \cos\left(\pi\left(v - D(1+\cos\phi)\right)\right)}{v - D(1+\cos\phi)} \\
&= \beta_r D \sin\phi \frac{\partial \mathbf{g}_r[v]}{\partial \alpha}, \\
\frac{\partial \mathbf{g}_i[v]}{\partial \phi} &= \beta_i D \sin\phi \frac{\operatorname{sinc}\left(v - D(1+\cos\phi)\right) - \cos\left(\pi\left(v - D(1+\cos\phi)\right)\right)}{v - D(1+\cos\phi)} \\
&= \beta_i D \sin\phi \frac{\partial \mathbf{g}_i[v]}{\partial \alpha}, \\
\frac{\partial \mathbf{g}_r[v]}{\partial \beta_r} &= \frac{\partial \mathbf{g}_i[v]}{\partial \beta_i} \\
&= \operatorname{sinc}\left(v - D(1+\cos\phi)\right), \\
\frac{\partial \mathbf{g}_r[v]}{\partial \beta_i} &= \frac{\partial \mathbf{g}_i[v]}{\partial \beta_r} \\
&= 0.
\end{aligned} \tag{4.27}$$

Therefore, the corresponding Fisher information matrix will be

$$\mathbf{F} = \frac{2}{\sigma^2}\begin{bmatrix} \|\beta\|^2 D^2(\sin^2\phi)\Phi & \beta_r D(\sin\phi)\Omega & \beta_i D(\sin\phi)\Omega \\ \beta_r D(\sin\phi)\Omega & \Sigma & 0 \\ \beta_i D(\sin\phi)\Omega & 0 & \Sigma \end{bmatrix}. \tag{4.28}$$

Based on the block matrix inversion theorem, we can get the diagonal elements of $\mathbf{F}^{-1}$ as

$$\mathbf{F}^{-1}[1,1] = \frac{\sigma^2}{2\|\beta\|^2 D^2(\sin^2\phi)(\Phi - \frac{\Omega^2}{\Sigma})}, \tag{4.29}$$

$$\mathbf{F}^{-1}[2,2] = \frac{\sigma^2\left(\|\beta\|^2\Phi\Sigma - \beta_i^2\Omega^2\right)}{2\Sigma\left(\|\beta\|^2\Phi\Sigma - \|\beta\|^2\Omega^2\right)}, \tag{4.30}$$

$$\mathbf{F}^{-1}[3,3] = \frac{\sigma^2\left(\|\beta\|^2\Phi\Sigma - \beta_r^2\Omega^2\right)}{\Sigma\left(\|\beta\|^2\Phi\Sigma - \|\beta\|^2\Omega^2\right)}, \tag{4.31}$$

where the last two terms are the same as in the previous case of estimating spatial signature. We can also see the CRLB varies with the nonlinear function '$\sin\phi$', which is similar to the result of ULA scenario. Intuitively, the estimation accuracy will be better when the incoming signal arrives at the position around the center beam while the estimation accuracy will be worse when it falls into the side beams. Same as in the ULA case, since the estimation accuracy is not stable, tracking the spatial signature is a better choice rather than the DoA itself by applying traditional array signal processing techniques.

4.5 Simulation results and outcome analysis

In this section, the numerical results for this single-user data-aided channel tracking system is presented from different perspectives, which is also the replayed results of [1] to verify the basis of this data-aided channel tracking system. The main motivation of this chapter is to give the closed-form expression of the CRLBs which are missing in [1], and next chapter is an extensional work of [1] with multi-user scenario. Therefore, the reproduction is very important to validate our understanding on the basis of this thesis.

The setup of this replay is consistent with the parameters defaulted in [1]. Here we still assume that the symbol duration is 10^{-6} second and the path loss factor is chosen to be 2 due to LoS only scenario. In our simulation, we assume the speed of the vehicle is 360 km per hour, which is typical for the very-high-speed railway system, and the radius of the cell is 500 meters. For initial channel estimation, if the length of training pilot is chosen to be 100. As for the parameters of the DLA, it really depends on how we choose the physical size of the EM lens. For example, the focal length is chosen to be 1 m, the number of antenna

elements is limited to be $M = 41$, and the number of active antenna elements of the user is $V = 3$. That is to say, the same orthogonal pilot has to be repeatedly transmitted 34 times to perform only once initial channel estimation. This will consume enormous resources when the cell radius is small, or equivalently handover is frequently happened, where the initial channel training is required right after the handover operation. In an attempt to reproduce the results in Figure 8 in [1], we start with $M = 128$, $N = 1000$, and $V = 3$ in our simulation, and the simulation result is shown in Figure 4.2. On the other hand, Figure 4.3 is taken from the [1] to show the original result. As we can see from our simulation result, the MSE of the tracking error approaches the theoretical CRLB. Compared with the red line in Figure 4.3, also with $M = 128$, the magnitudes of MSE are almost the same with our numerical results. At this point, we are confident with the data-aided channel tracking process.

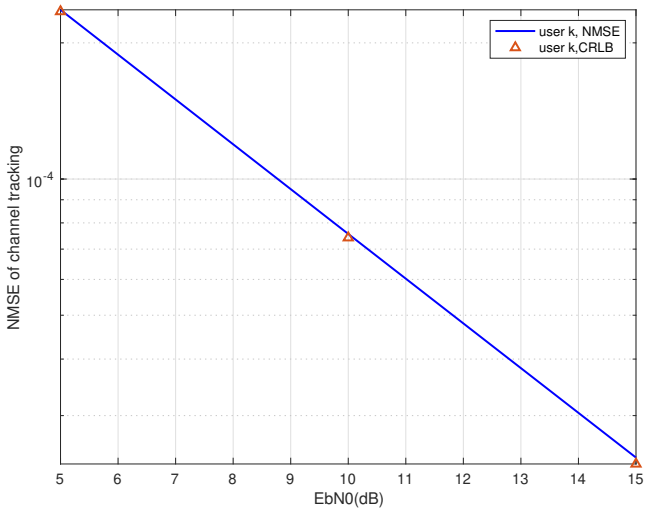

Figure 4.2: Replayed result of Figure 8 in [1]

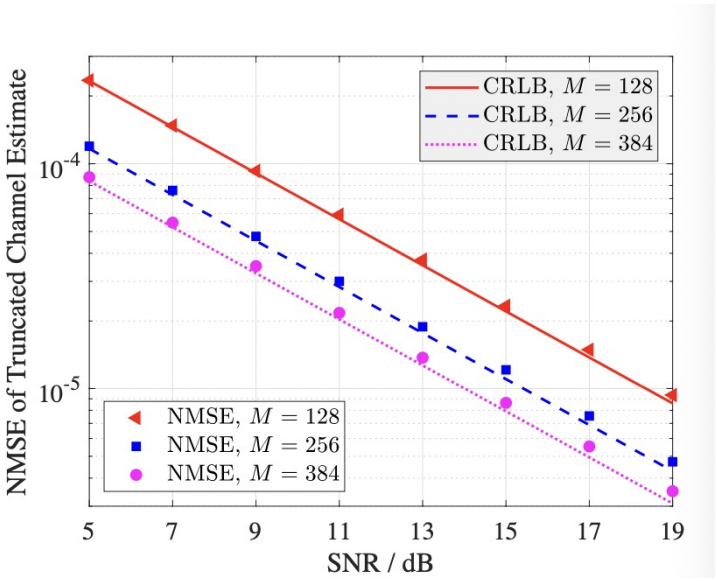

Figure 4.3: Preliminary channel tracking error results [1]

Next, we will elaborate the impacts on the channel tracking performance from different aspects, e.g., the number of antennas M, the number of active antennas for truncated channel V and the block length N. First of all, the impact of the number of antennas on NMSE is presented in Figure 4.4, where $M \in \{64, 128, 192, 256\}$. Apparently, more antennas help to concentrate the beam and thus improve the energy efficiency. However, due to the truncated channel vector model, we lost nearly 20% energy while doing data detection, which degrades the SER performance. As a result, we can see that there is a gap between the simulation result and CRLB at low SNR regime when power gain is not enough to compensate the truncated channel energy loss. This phenomenon is much more severe for small-scale MIMO system, which can be seen in Figure 4.5. At low SNR regime, where $M \in \{16, 32, 48, 64\}$, the gap between the evaluations and the theoretical CRLBs is large. As the number of antennas increases, the power gain rises, thus achieving better

channel tracking accuracy.

However, it should be noted that this conclusion is based on the simulation results by taking beamforming gain into account. As we have discussed in the previous chapter, the DLA performs as an analog beamformer due to its optical characteristics. As a result, the SNR in this chapter is defined as the output SNR after taking the beamforming gain into account, thus adding much higher noise than the counterpart of ULA case. If we make it as the input SNR before beamforming, the performance is exactly the same with Figure 3.5. The reason is that DLA is the only approximation of DFT operation of the channel response of ULA, without adding any new feature, which results in the same channel tracking performance.

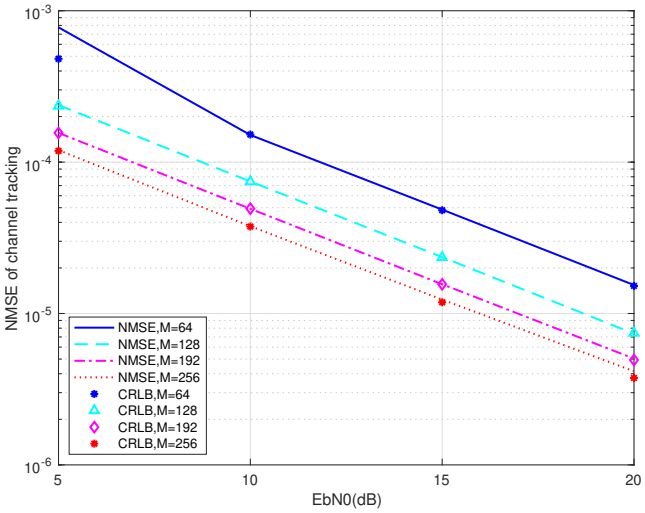

Figure 4.4: NMSE and CRLB of tracking error for large values of M.

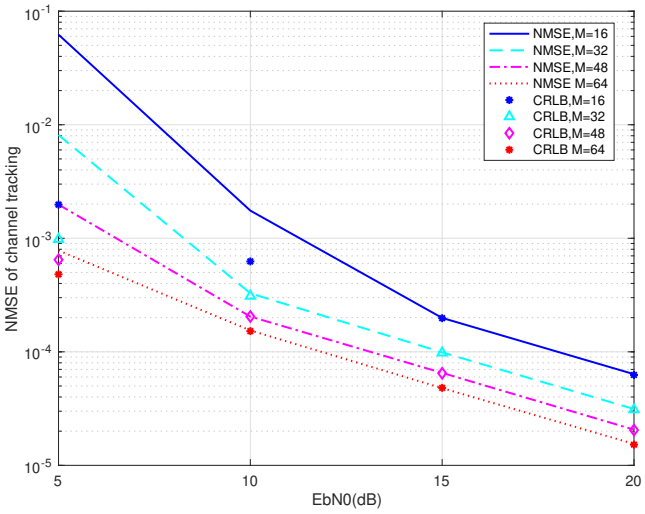

Figure 4.5: NMSE and CRLB of tracking error for small values of M.

In addition, we will show the impact of the number of active antennas on the system performance. As for V, we need to carefully choose. On one side, it can not be large because the purpose to introduce truncated channel vector is to save cost. If $V \times K \geq M$, this concept losses its significance. On the other side, it can not be too small neither to make *transition* process happen [1]. As proven in [1], at least three antennas should be activated to achieve comparable performance with respect to that with M active antennas. As we have seen in Figure 3.2, the two antennas in the main lobe contain more than 80% of energy. That is to say, once we increase the number of active antenna number, the amount of energy we can collect increases, thus upgrading the system performance. With respect to DoA with 0.4π, Figure 4.6 shows the channel tracking results at the 100-th time slot with V varying in $\{3, 6, 9\}$, where the simulation results approach the theoretical CRLBs for all cases as expected. However, the accuracy improvement is almost negligible every

time we double the active antennas number V.

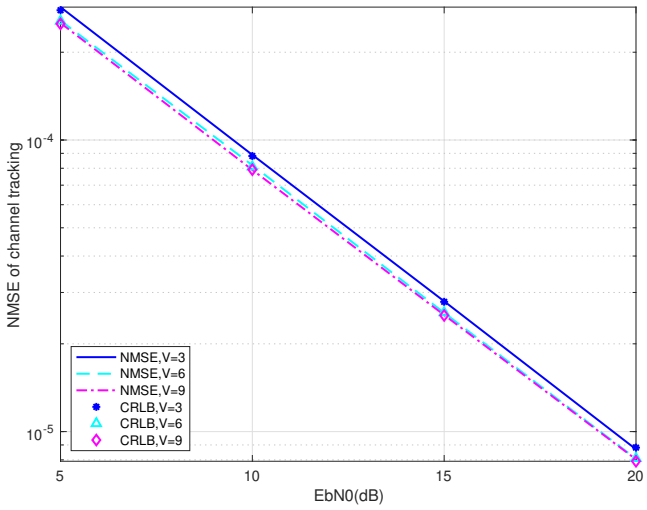

Figure 4.6: NMSE and CRLB of tracking error with different V.

Moreover, this impact will be most negligible when there is only one antenna element in the main lobe which contains all the energy while the others happen to be at the samples with 0 value of the sinc function, which is shown in Figure 4.8. Therefore, with respect to DoA with $1/3\pi$, Figure 4.7 shows the channel tracking performance at the 100-th time slot with V varying in $\{3, 6, 9\}$, where the simulation results are approaching the theoretical CRLBs as well. And the accuracy improvement is even more negligible than the previous case. To sum up, the impact of the number of active antennas V on the system performance is negligible since most energy is collected in the main lobe.

Figure 4.9 presents the impact of the block length N on the system accuracy performance. Every time we double the simulated values of N, the NMSE of channel estimation will be reduced by a factor of two, which is consistent with the CRLB analysis in the

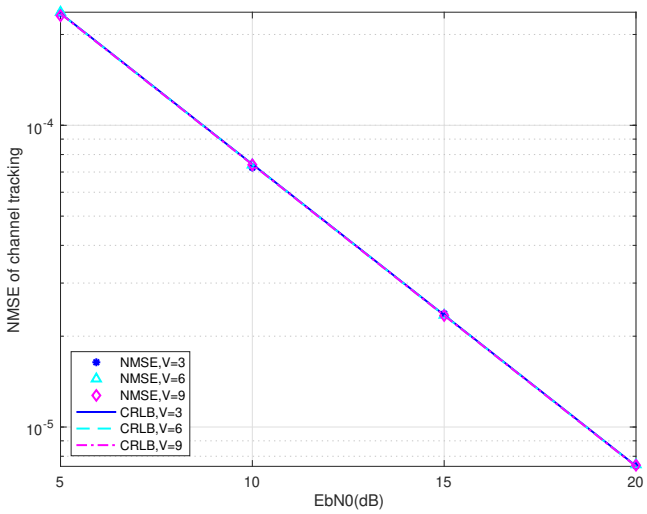

Figure 4.7: NMSE and CRLB of tracking error with different N.

previous section. However, choosing an appropriate block length N is a trade-off. On the positive side, longer data block can help to enhance the channel tracking accuracy by suppress noise. On the negative side, we are using the channel estimates from the previous time slot for the data detection of the current block, since a high temporal correlation is assumed. Therefore, longer data blocks may lead to larger channel variations and data detection errors, which may propagate and deteriorate the whole system performance.

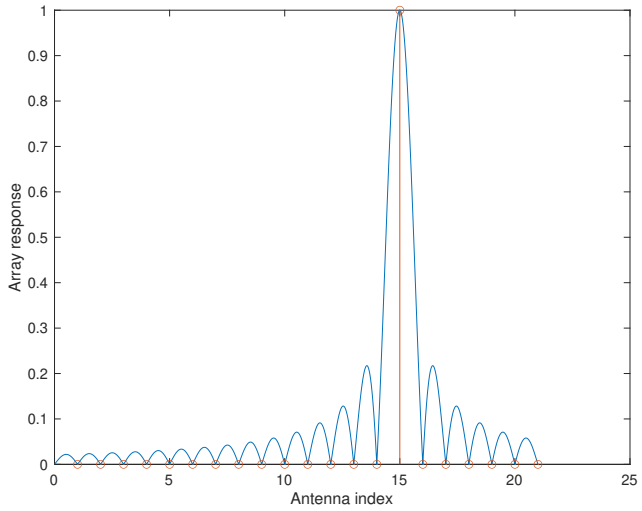

Figure 4.8: Channel response with DoA $\frac{1}{3}\pi$.

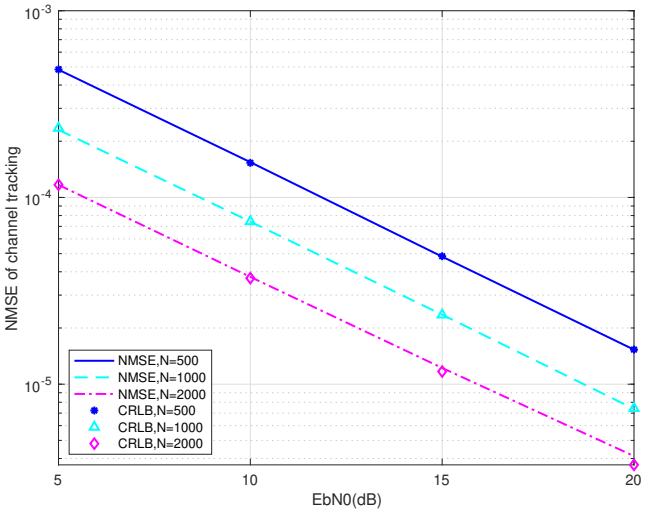

Figure 4.9: NMSE and CRLB of tracking error with different N.

4.6 Summary

In this chapter, a DLA-based data-aided channel tracking framework is presented for mmWave massive MIMO system with limited RF chains. Firstly the DLA array response has been comprehensively studied and then the data-aided channel tracking method is adopted from [1]. The contribution is that we give the closed-form expression of CRLBs with respect to spatial signature estimation and conventional DoA estimation respectively. This CRLB analysis also proves that spatial signature estimation is the better choice than DoA estimation for channel prediction solutions. From the simulation results, we can clearly see that the Newton method based data-aided channel tracking mechanism closely approaches the CRLB. In the next chapter, we will extend this work with multi-user scenario by dealing with co-channel issues.

Chapter 5

Multi-user data-aided channel tracking framework

The Discrete Lens Antenna (DLA-based) communication system, proposed in [10] with impressive experimental result, becomes one promising solution for mmWave communication to compensate for the huge attenuation with limited RF chains. According to the optical characteristics of the electromagnetic (EM) lens, DLA performs as a analog beamformer by approximating a DFT operation to transform the signal to spatial domain, thus exploring the spatial sparsity in mmWave communication. Specifically, since most energy will be only concentrated on a small number of antenna elements, it is reasonable to reduce cost and complexity by using limited RF chains. To this end, channel tracking method is promising instead of traditional channel estimation, which consumes a large amount of resources by repeatedly transmitting the same pilot. The authors in [1] explored a similar DLA-based channel tracking process for single-user scenario, while this chapter illustrates

the additional problems and solutions in multi-user scenario.

In this Chapter, the far-field DLA-based channel model is firstly explored, and the truncated counterpart is then introduced. Based on that, spatial resolution issue with DLA is discussed to make inter-user interference (ISI) more serious when multiple users are within the same wide beam. To overcome this ISI issue, we propose an advanced data-aided channel tracking method to keep estimating and updating the individual channel vector when multiple users are not well separated. Since we demonstrate that the SER performance in data detection is much more vulnerable than channel estimation error, we proposed an estimated SER metric for scheduling decision to deal with ISI in multi-user scenario.

5.1 DLA-based system model

5.1.1 Channel model

Without loss of generality, we assume that, in each cell, the base station (BS) with DLA serves up to K mobile users simultaneously, each fitted with one antenna. As each EM lens is with normalized dimension of D, the number of antenna elements on the focal arc will be $M = 1 + 2\lfloor D \rfloor$. Furthermore, in V2I (vehicle-to-infrastructure) mmWave communication, LoS path dominates due to the high signal propagation loss. Another critical assumption here is that the Doppler effect due to the high mobility is eliminated on the RF end via Phase-locked Loops (PLLs) [1]. Therefore, by adopting the far-field flat channel model, the

[1] The Doppler effect is out of the scope of this entire thesis.

beamspace channel vector $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ of the k-th user is given as

$$\mathbf{h}_k = \beta_k \mathbf{a}\left(\phi_k\right), \tag{5.1}$$

$$\mathbf{h}_k[m] = \beta_k \mathrm{sinc}(m - \alpha_k), \forall m \in [0, M-1], \tag{5.2}$$

where $\mathbf{a}(\phi_k)$ is the array response as shown in Theorem 2 with $\alpha_k = D(1+\cos\phi_k)$, β_k is the equivalent complex gain, and ϕ_k is the DoA of the received signal coming from the k-th mobile terminal.

5.2 Problem statement

5.2.1 Motivation

It has been claimed that DLA-based mmWave communication system with limited RF chains can reduce cost and complexity to a great extend. However, the common problem with any limited RF chains system is the high overhead of channel estimation. Recall that V antennas are activated for each user, the total number of RF chains is thus KV, or equivalently only KV antenna elements need to be scanned during any single time slot. As a result, to scan the entire channel vector, each orthogonal pilot sequence is required to be repeatedly transmitted for $\lceil M/(KV)\rceil$ time slots, which consumes enormous amount of resources. To this end, channel tracking becomes a better choice (e.g. data-aided channel tracking [1] and fast tracking with low pilot overhead [59]). As we discussed previously, [59] gives reasonable estimation error performance without considering data detection part, whose phase error is not acceptable for uplink data detection or downlink

data transmission. In comparison, [1] provides a sophisticated data-aided approach from system-level perspective. In [1], since channel tracking process highly replies on the data detection results, it is crucial to make data detection trustable. To this end, even with powerful channel coding theory, IUI is still an inevitable issue to deal with in multi-user scenario. To avoid this issue, [1] assumes that all the users are well separated in angle domain by some sophisticated scheduling method, thus exploring the single-user scenario, where we can disclose the parametric impact on the channel tracking accuracy clearly. However, to make this scheduling decision is still an on-going research topic. As a result, in this chapter, we will investigate the multi-user scenario by illustrating IUI effect in DLA-based communication system, proposing the counterpart channel tracking method for multi-user scenario and establishing a new metric for scheduling approach.

5.2.2 Spatial resolution

For conventional ULAs, antenna elements are spaced with a pitch of half wavelength where the angle domain is equally divided by the number of antennas M. However, for DLA arrays, the so-called critical spacing means that the antenna elements are located on the focal arc, based on the fact that $\tilde{\theta}_m \triangleq \sin\theta_m$, in the interval $[-1, 1]$. In other words, only the projection of the antenna array is equally spaced since *sin* function is nonlinear, which leads to the fact that the placements of the central antennas are dense while those of the side antennas are sparse. This fact also results in great spatial resolution degradation for the side beams because the beam width is much wider than that of central beams. Furthermore, the key reason for spatial resolution degradation is the critical assumption $F \gg D_y, D_z$ in [2], which is also in the derivation of the DLA response in Appendix A. In

that case, the size of EM lens highly affects the normalized dimension D, which therefore limits the total number of antenna elements M.

Intuitively, the spatial frequency resolution of DLA is defined inversely proportional to the quantity of D, which means that two users with DoAs of ϕ_i and ϕ_j $(\forall i \neq j)$ can be sufficiently separated only when $|\cos\phi_i - \cos\phi_j| \geq 1/D$ is satisfied [2]. However, D is not possible to be large. For the comparable sizes of ULA and DLA arrays, the number of antenna elements of the latter one can only be as large as 20% of that of ULA array. For example, assume the physical size of both ULA and DLA is 1 m, and the wavelength is 1 cm which is typical for 30 GHz, then the ULA could consist of 100 antennas. However, for the DLA, the focal length determines the physical size of the DLA and the critical assumption, $F \gg D_y, D_z$, limits the size of the EM to be 10 cm. Therefore, the counterpart number of the DLA could only be 21 based on the fact that $M = 1 + 2\lfloor D \rfloor$. This results in a significant degradation of the spatial resolution of DLA array, especially for the two widest side beams.

5.2.3 Multi-user co-channel issue

In multi-user scenario, IUI has to be inevitably treated as a significant issue especially for the side beams, where main lobes of different users will be overlapped for a long time period. Figure 5.1 illustrates a practical scenario of two users with speeds of 150 km/h and 280 km/h respectively, in which case user 2 chases and passes user 1. In this setup, $M = 41$ and thus the normalized dimension of EM lens $D = 20$, which are reasonable. Figure 5.2 shows how their DoAs vary along the time indices, and it is obvious that they intersect while overtaking. According to the definition of spatial resolution of DLA, as

shown in Figure 5.3, the angle variations can be represented in their trigonometric form, and red line indicates the time indices when two users are unresolvable in spatial domain, or equivalently $|\cos\phi_1 - \cos\phi_2| \leq 0.05$. Figure 5.4 shows the channel correlation between user 1 and user 2, and the main lobe indicates the duration of the main lobes of their channel responses being overlapped. This co-channel issue can last up to hundreds of time slots, as shown in Figure 5.3, and is even worse for the wider beams around the two ends, which will cause inevitable channel tracking error and enormous detection error for the whole system. Therefore, this extensional work of [1] focuses on the co-channel multi-user scenario.

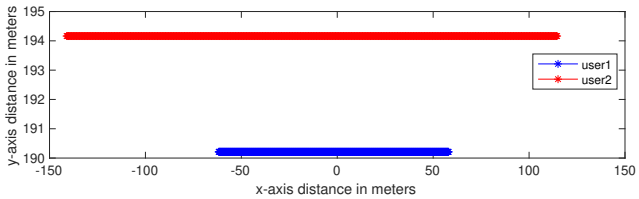

Figure 5.1: Trajectory setup for overtaking scenario with two users.

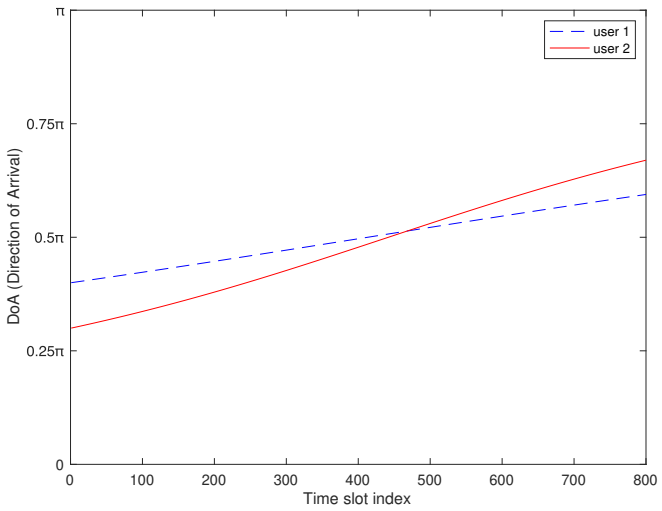

Figure 5.2: Linear DoAs (ϕ_k) variation versus time indices.

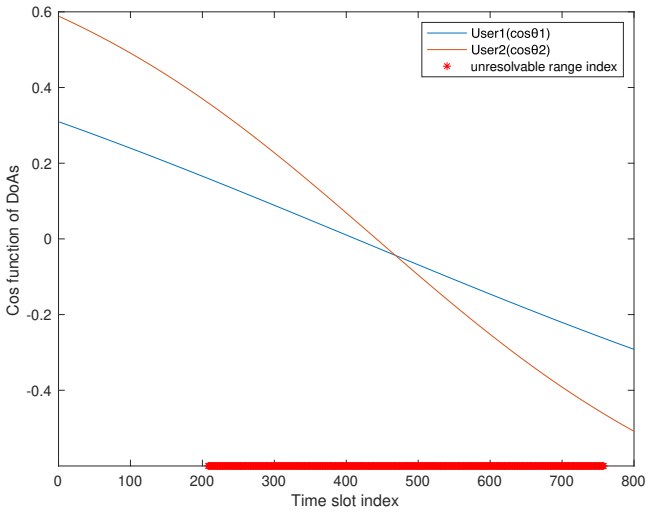

Figure 5.3: Nonlinear spatial resolution($\cos\phi_k$) variation versus time indices. Red points indicate the time period when two users are being interfered.

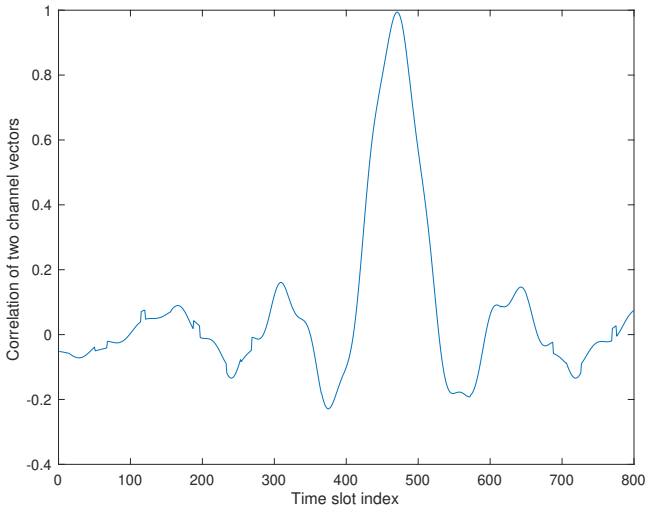

Figure 5.4: Interference also resembling 'sinc' waveform.

5.3 Data-aided channel tracking for multi-user scenario

As data-aided channel tracking methods save a great amount of resources and provide trustable estimation result for both uplink and downlink, we will follow the similar method via data-aided channel tracking with various multiple access strategies to deal with co-channel multi-user channel tracking issue. Without loss of generality, we assume time-division duplex (TDD) to explore the channel reciprocity, and only uplink channel estimation is employed. In the conventional mmWave systems with channel estimation, the three standard phases in each time slot include processing, downlink beamforming and uplink data detection. Pilot training is typically done during the processing phase, and this estimated CSI will be used for further downlink beamforming and uplink detection in the same time slot. As we have discussed, with limited RF chains, to estimate the en-

tire channel vector requires the same pilot to be repeatedly transmitted $\lceil M/(KV) \rceil$ times, which consumes a large amount of resources. Therefore, this work will follow the similar data-aided channel tracking method as in [1], where the data-aided channel estimates in the current time slot is used for downlink beamforming and uplink data detection in the next time slot. The authors in [1] focus on single user scenario based on the assumption that all users are well separated in angle domain, which is not always the case without well-performed scheduling methods. However, the metric of the scheduling approach may depend on the estimation performance. To this end, this work will focus on multi-user scenario with co-channel issue. Through this chapter, to simplify the notations, we assume there are two users in the single cell with DoAs of ϕ_1 and ϕ_2.

5.3.1 Initial channel estimation

In the initial channel estimation phase, conventional pilot training process is conducted with orthogonal pilot sequences, thus ignoring inter-user interference. However, due to the limited RF chains, it has to take at least $\lceil M/(KV) \rceil$ time slots to scan the entire antenna array since only V antenna elements can be activated during each time slot. After repeatedly transmitting the orthogonal pilots from all users, the uplink received signal by BS at 0-th time slot $\mathbf{Y}_0 \in \mathbb{C}^{M \times \tau}$ is given as

$$\mathbf{Y}_0 = \sum_{k=1}^{K} \mathbf{h}_{k,0} \mathbf{p}_k^T + \mathbf{N}_0, \ \forall K = 2, \tag{5.3}$$

where $\mathbf{p}_k \in \mathbb{C}^{\tau \times 1}$ denotes the orthogonal pilot sequence of the k-th user with length of τ, $\mathbf{h}_{k,0} \in \mathbb{C}^{M \times 1}$ means the channel vector of the k-th user at 0-th time slot, and $\mathbf{N}_0 \in \mathbb{C}^{M \times \tau}$

is additive Gaussian noise, in which each element is i.i.d. zero-mean circularly symmetric complex Gaussian (CSCG) random variables with variance σ_n^2. By applying simple least square estimator, the estimate of the k-th user will be

$$\hat{\mathbf{h}}_{k,0} = \frac{\mathbf{Y}_0 \mathbf{p}_k^*}{\|\mathbf{p}_k\|^2} = \frac{\mathbf{h}_k \mathbf{p}_k^T \mathbf{p}_k^*}{\|\mathbf{p}_k\|^2} + \sum_{k' \neq k} \frac{\mathbf{h}_{k'} \mathbf{p}_{k'}^T \mathbf{p}_k^*}{\|\mathbf{p}_k\|^2} + \frac{\mathbf{N}_0 \mathbf{p}_k^*}{\|\mathbf{p}_k\|^2}. \tag{5.4}$$

With the help of orthogonal pilots $\mathbf{p}_1 \in \mathbb{C}^{\tau \times 1}$ and $\mathbf{p}_2 \in \mathbb{C}^{\tau \times 1}$, the channel estimates of two users $\hat{\mathbf{h}}_{1,0} \in \mathbb{C}^{M \times 1}$ and $\hat{\mathbf{h}}_{2,0} \in \mathbb{C}^{M \times 1}$ are given as

$$\hat{\mathbf{h}}_{1,0} = \mathbf{h}_1 + \frac{\mathbf{N}_0 \mathbf{p}_1^*}{\|\mathbf{p}_1\|^2}, \quad \hat{\mathbf{h}}_{2,0} = \mathbf{h}_2 + \frac{\mathbf{N}_0 \mathbf{p}_2^*}{\|\mathbf{p}_2\|^2}. \tag{5.5}$$

According to the aforementioned fact, most energy concentrates on V antenna elements, thus enabling jointly truncated channel vector with length of KV. Algorithm 2 elaborates the method to find the indices vector of active antenna element $\mathbf{I}_n \in \mathbb{R}^{KV \times 1}$ and the resulting jointly truncated channel vector with multiple users at the n-th time slot. By Algorithm 2, the jointly truncated channel vector for each user can be obtained as

$$\hat{\mathbf{h}}_{1,t,0} = \hat{\mathbf{h}}_{1,0}[\mathbf{I}_0], \quad \hat{\mathbf{h}}_{2,t,0} = \hat{\mathbf{h}}_{2,0}[\mathbf{I}_0]. \tag{5.6}$$

It should be noted that these active antennas will be used for DL transmission and UL reception in the next time slot. Therefore, these initial estimates will be updated as the estimates of the n-th time slot, which will be used in the $(n+1)$-th time slot to activate channel tracking process.

Algorithm 2 Multi-user jointly channel truncating algorithm.

Identify $\mathbf{I}_{1,n} = [v_{1,n} : v_{1,n} + V - 1]$ and $\mathbf{I}_{2,n} = [v_{2,n} : v_{2,n} + V - 1]$ via Algorithm 1, and define corresponding indices sets as $I_{1,n} = \{\mathbf{I}_{1,n}[a] | a = 0, 1, ..., V - 1\}$ and $I_{2,n} = \{\mathbf{I}_{2,n}[b] | b = 0, 1, ..., V - 1\}$
Jointly find the entire indices of active antennas $\mathbf{I}_n$
if $|I_{1,n} \cap I_{2,n}| = 0$ **then**
  $\mathbf{I}_n$ being a contamination of two consecutive index segments as $\mathbf{I}_n = [\min(v_{1,n}, v_{2,n}) : \min(v_{1,n}, v_{2,n}) + V - 1, \max(v_{1,n}, v_{2,n}) : \max(v_{1,n}, v_{2,n}) + V - 1]$
else if $|I_{1,n} \cap I_{2,n}| = i \neq 0$ **then**
  $\mathbf{I}_n$ being a consecutive index segment with number of $2V - i$ active antennas
  let $\mathbf{I}' = [\min(v_{1,n}, v_{2,n}) : \min(v_{1,n}, v_{2,n}) + 2V - 1 - i]$ be the ordinal indices
  if i is even **then**
    symmetrically pick $i/2$ indices away from both ends of $\mathbf{I}'$, and check whether the new ends reach the starting and ending index of $m \in [0, M-1]$, similar to Algorithm 1 to get starting index v_n of the jointly truncated channel vector
  else if i is odd **then**
    symmetrically pick $(i-1)/2$ indices away from both ends of $\mathbf{I}'$, and pick additional one sample from either side with stronger energy;
    check whether the new ends reach the starting and ending index of $m \in [0, M-1]$, similar to Algorithm 1 to get starting index v_n
  end if
end if
return $\mathbf{I}_n = [v_n : v_n + 2V - 1]$ and truncated channel vector of each user as $\hat{\mathbf{h}}_{1,t,n} = \hat{\mathbf{h}}_1[\mathbf{I}_n]$ and $\hat{\mathbf{h}}_{2,t,n} = \hat{\mathbf{h}}_{2,t}[\mathbf{I}_n]$;

5.3.2 Data-aided channel estimation

Assume $\mathbf{I}_n = [v_n : v_n + 2V - 1]$, $\hat{\mathbf{h}}_{1,t,n}$ and $\hat{\mathbf{h}}_{2,t,n}$ obtained from the n-th time slot. At the $(n+1)$-th time slot, the uplink received signal $\mathbf{Y}_{n+1} \in \mathbb{C}^{2V \times L}$ is the superposition of all the users

$$\mathbf{Y}_{n+1} = \sum_{k=0}^{K} \mathbf{h}_{k,d,n+1}\mathbf{x}_{k,n+1} + +\mathbf{N}_{n+1} = \mathbf{h}_{1,d,n+1}\mathbf{x}_{1,n+1}^T + \mathbf{h}_{2,d,n+1}\mathbf{x}_{2,n+1}^T + \mathbf{N}_{n+1}, \tag{5.7}$$

where $\mathbf{h}_{k,d,n+1} = \mathbf{h}_{k,n+1}[\mathbf{I}_n]$ denotes the truncated true channel vector of the k-th user at the $(n+1)$-th time slot with the active antenna indices updated from the last time slot, $\mathbf{N}_{n+1} \in \mathbb{C}^{2V \times L}$ is additive Gaussian noise, in which each element is i.i.d. zero-mean circularly symmetric complex Gaussian (CSCG) random variables with variance σ_n^2, and L is the length of the transmitted uplink data $\mathbf{x}_{k,n+1}$ during a single time slot. It is widely accepted that beamspace channel vectors between any two adjacent time slots are high correlated since multipath effect in mmWave communications is not severe. This allows us to utilize the channel estimates $\hat{\mathbf{h}}_{k,t,n}$ from the previous time slot as the approximates of $\mathbf{h}_{k,d,n+1}$ to do data detection in the $(n+1)$-th time slot. In the following phases, we will take user 1 as an example, and similar process can be directly adapted to user 2. Therefore, the recovered signal $\tilde{x}_{n+1} \in \mathbb{C}^{L \times 1}$ via MF detector is given by

$$\tilde{\mathbf{x}}_{1,n+1} = \frac{\mathbf{Y}_{n+1}^T \hat{\mathbf{h}}_{1,t,n}^*}{\|\hat{\mathbf{h}}_{1,t,n}\|^2}. \tag{5.8}$$

Since data detection is only processed within each $(n+1)$-th time slot, and $\hat{\mathbf{h}}_{k,d,n+1}$ will be updated as $\hat{\mathbf{h}}_{k,t,n}$ at the end of each time slot, it allows us to ignore the subscript of n and $n+1$ here to achieve notational simplicity. Then equation (5.8) can be rewritten as

$$\tilde{\mathbf{x}}_1 = \frac{\mathbf{Y}^T \hat{\mathbf{h}}_{1,t}^*}{\|\hat{\mathbf{h}}_{1,t}\|^2} = \frac{\mathbf{x}_1 \mathbf{h}_{1,d}^T \hat{\mathbf{h}}_{1,t}^*}{\|\hat{\mathbf{h}}_{1,t}\|^2} + \frac{\mathbf{x}_2 \mathbf{h}_{2,d}^T \hat{\mathbf{h}}_{1,t}^*}{\|\hat{\mathbf{h}}_{1,t}\|^2} + \mathbf{w}_1, \tag{5.9}$$

where $\mathbf{w}_1 = \frac{\mathbf{N}^T \hat{\mathbf{h}}_{1,t}^*}{\|\hat{\mathbf{h}}_{1,t}\|^2}$, and $\tilde{\mathbf{x}}$ is subject to detection error resulting from estimation error, inter-user interference and noise. By employing powerful channel coding methods (e.g. Turbo codes and LDPC) with moderate interference, we are able to obtain a trustworthy

estimate $\hat{\mathbf{x}}_1$ of the original data vector $\mathbf{x}_1$. At this point, the data-aided channel estimates will be given by

$$\hat{\mathbf{h}}_{1,d} = \frac{Y\hat{\mathbf{x}}_1^*}{||\hat{\mathbf{x}}_1||^2} = \frac{\left(\mathbf{h}_{1,d}\mathbf{x}_1 + \mathbf{h}_{2,d}\mathbf{x}_2 + \mathbf{N}\right)\hat{\mathbf{x}}_1^*}{||\hat{\mathbf{x}}_1||^2} \approx \hat{\mathbf{h}}_{1,d} + \frac{\mathbf{h}_{2,d}\mathbf{x}_2\hat{\mathbf{x}}_1^*}{||\hat{\mathbf{x}}_1||^2} + \frac{\mathbf{N}\hat{\mathbf{x}}_1^*}{||\hat{\mathbf{x}}_1||^2}. \tag{5.10}$$

As long as the length of data sequence L is large, the second term is negligible due to asymptotical orthogonality. Therefore, it should be noted that the estimation accuracy is limited by the detection error, or Symbol Error Rate (SER) performance, which will be discussed in the later part. Based on the estimation of the truncated beamspace channel vector $\hat{\mathbf{h}}_{1,d}$ of user 1, we need to recover the entire channel vector $\hat{\mathbf{h}}_1$ via Newton method to facilitate the *transition* process which will be illustrated in the following section.

5.3.3 *Transition* phase

In this section, we will present a Newton method solution for this channel tracking problem, which will be utilized by each user independently to obtain the complex gains $\hat{\beta}_1$, $\hat{\beta}_2$ and corresponding spatial signatures $\hat{\alpha}_1$, $\hat{\alpha}_2$. Since the indices vector $\mathbf{I}_n$ in the current time slot will be used to obtain the truncated channel vector in the next time slot to perform downlink transmission and uplink detection, it is critical to adaptively change the active antenna indices to guarantee acceptable performance [1]. To this end, we need to approximately reconstruct the estimates of the entire channel vectors of each user at the n-th time slot as

$$\begin{aligned} \hat{\mathbf{h}}_{1,n}[m] &= \hat{\beta}_1 \mathrm{sinc}(m - \hat{\alpha}_1), \\ \hat{\mathbf{h}}_{2,n}[m] &= \hat{\beta}_2 \mathrm{sinc}(m - \hat{\alpha}_2), \ \ \forall m \in [0, M-1]. \end{aligned} \tag{5.11}$$

To obtain a close approximation of this optimal solution, we will apply a low complexity LS estimation via iterative Newton algorithm. First of all, it is clear that there are two parameters to be estimated $\boldsymbol{\theta} \triangleq \left[\alpha_{k,n+1}, \beta_{k,n+1}\right]^T$. To simplify the notations, we define the following without subscript of user index

$$\mathbf{h}_r = \mathcal{R}\left\{\mathbf{h}_{d,n+1}\right\}, \mathbf{h}_i = \mathcal{F}\left\{\mathbf{h}_{d,n+1}\right\}, \tag{5.12}$$

$$\hat{\mathbf{h}}_r = \mathcal{R}\left\{\hat{\mathbf{h}}_{d,n+1}\right\}, \hat{\mathbf{h}}_i = \mathcal{F}\left\{\hat{\mathbf{h}}_{d,n+1}\right\}, \tag{5.13}$$

$$\beta_r = \mathcal{R}\left\{\beta_{n+1}\right\}, \beta_i = \mathcal{F}\left\{\beta_{n+1}\right\}, \tag{5.14}$$

$$\hat{\beta}_r = \mathcal{R}\left\{\hat{\beta}_{n+1}\right\}, \hat{\beta}_i = \mathcal{F}\left\{\hat{\beta}_{n+1}\right\}. \tag{5.15}$$

To this end, we have the MLE of θ as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\hat{\mathbf{h}}_r - \mathbf{h}_r\|^2 + \|\hat{\mathbf{h}}_i - \mathbf{h}_i\|^2 = \arg\min_{\boldsymbol{\theta}} \|\hat{\mathbf{h}}_s - \mathbf{h}_s\|^2, \tag{5.16}$$

where $\hat{\mathbf{h}}_s$ and $\mathbf{h}_s$ are defined as

$$\hat{\mathbf{h}}_s = \left[\hat{\mathbf{h}}_r^T, \hat{\mathbf{h}}_i^T\right]^T, \mathbf{h}_s = \left[\mathbf{h}_r^T, \mathbf{h}_i^T\right]^T. \tag{5.17}$$

By using Newton method, we can get the estimate of $\boldsymbol{\theta}$ at the $(j+1)$-th iteration, namely $\hat{\boldsymbol{\theta}}^{(j+1)}$, based on the previous estimate at the j-th iteration $\hat{\boldsymbol{\theta}}^{(j)}$ as

$$\hat{\boldsymbol{\theta}}^{(j+1)} = \left(\mathbf{S}^H\mathbf{S}\right)^{-1}\mathbf{S}^H\left(\hat{\mathbf{h}}_s - \mathbf{s}\left(\hat{\boldsymbol{\theta}}^{(j)}\right)\right) + \hat{\boldsymbol{\theta}}^{(j)}, \tag{5.18}$$

where $\mathbf{s}(\boldsymbol{\theta}) = \left[\mathbf{s}_r^T, \mathbf{s}_i^T\right]^T$ is given as

$$\mathbf{s}_r(\boldsymbol{\theta})[v] = \beta_r \text{sinc}\left(v_{0,n} + v - \alpha_{n+1}\right), \tag{5.19}$$

$$\mathbf{s}_i(\boldsymbol{\theta})[v] = \beta_i \text{sinc}\left(v_{0,n} + v - \alpha_{n+1}\right), \ \forall v \in [0, 2V-1], \tag{5.20}$$

and $\mathbf{S} \in \mathbb{R}^{2V\times 3}$ is the Jacobi matrix at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(j)}$:

$$\mathbf{S} = \left[\frac{\partial \mathbf{h}_r^T}{\partial \boldsymbol{\theta}}, \frac{\partial \mathbf{h}_i^T}{\partial \boldsymbol{\theta}}\right]^T \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(j)}}. \tag{5.21}$$

By doing this, the final estimate of $\hat{\boldsymbol{\theta}} = \left[\hat{\alpha}_{n+1}, \hat{\beta}_r, \hat{\beta}_i\right]$ can be obtained. Then it is clear to backtrack the MLE of β_{n+1} with its real and imaginary parts as $\hat{\beta}_{n+1} = \hat{\beta}_r + i\hat{\beta}_i$. To this end, based on the estimated $\hat{\alpha}_{n+1}$ and $\hat{\beta}_{n+1}$, it is straightforward to reconstruct the entire channel vector via channel model in section 5.1

$$\hat{\mathbf{h}}_{n+1}[m] = \hat{\beta}_{n+1} \text{sinc}\left(m - \hat{\alpha}_{n+1}\right). \tag{5.22}$$

This least square solution follows standard Newton method, which requires a carefully chosen initial point to guarantee its fast convergence. In this design, the estimate at the current time slot is chosen as the initial estimate in the Newton method based on the fact of high channel correlation between any two adjacent time slots, or equivalently $\alpha_{n+1} \approx \alpha_n$ and $\beta_{n+1} \approx \beta_n$. Therefore, the initial estimate for the first iteration in the $(n+1)$-th time slot is chosen based on the n-th time slot channel estimation

$$\hat{\boldsymbol{\theta}}^{(0)} = \left[\hat{\alpha}_n, \mathcal{R}\left\{\hat{\beta}_n\right\}, \mathcal{F}\left\{\hat{\beta}_n\right\}\right]. \tag{5.23}$$

After a small number of iterations, the complete channel vector can be derived based on the final result of $\hat{\boldsymbol{\theta}}$. At this point, it is necessary to re-select the antenna indices to be activated for the next time slot based on the new channel estimation result. The reason is that the DLA array response is sliding along the entire antenna array on the focal arc as the spatial signature changes, or equivalently as user moves inside the cell. However, with limited RF chains, not all the antenna elements are activated. If the activated antenna elements are not adaptively changed over time, it is impossible to receive the main lobe of the sinc function when the main lobe of the user is moving, thus disabling the truncated channel vector. Therefore, re-selection of the active antenna indices, based on the strongest channel response, is critical to guarantee the validation of the truncated channel vector with limited RF chains. Using Algorithm 2, the new active antenna indices of the current time slot can be obtained as $\mathbf{I}_{n+1} = [v_{0,n+1} : v_{0,n+1} + 2V - 1]$. As a result, the new jointly truncated channel vector will be

$$\hat{\mathbf{h}}_{k,t,n+1} = \hat{\mathbf{h}}_{k,n+1}[\mathbf{I}_{n+1}]. \tag{5.24}$$

For single-user scenario, this truncated channel estimation can be used for data transmission and detection in the next time slot based on the high correlation between two adjacent time slots, i.e., $\hat{\mathbf{h}}_{k,t,n} \approx \hat{\mathbf{h}}_{k,d,n+1}$. However, it might be a different story in multi-user scenario. Due to the inter-user interference, the data detection result in the current time slot may not be trustable enough to recover the complete channel estimation, thus causing severe channel estimation error, which will be propagated into the next time slot. Therefore, in this chapter, a new phase called *Update* is introduced, in which the SER performance

would be estimated based on the truncated channel estimation in the current time slot to determine if it could be updated for the next time slot to facilitate the whole tracking process. The details of the *Update* phase will be introduced in the following part.

5.3.4 *Update* phase

In this section, we will introduce the *Update* phase and the corresponding scheduling approach based on estimated SER metric. As we can see from the main body of this data-aided tracking process via equations (5.9) and (5.10), the majority of the channel tracking error comes from uplink detection error, which results from estimation error at the previous time slot, inter-user interference and noise at the current time slot. As shown in [1], in single-user scenario, noise dominates in channel tracking error, which guarantees the accuracy of the estimates as long as the tracking process is executed frequently enough. However, in multi-user scenario, inter-user interference can dominates in detection error when channel vectors of multiple users are spatially unresolvable. When users are well separated, inter-user interference is at noise level, which will not degrade tracking performance severely. However, when mobile users become too close to each other to be distinguished in angle domain, or equivalently the main lobes of these co-channel users are overlapped, as shown in Figure 5.4, IUI becomes severe in uplink data detection. Once the SER becomes worse beyond a threshold ε, the system will experience outage, thus the channel tracking process is crushed. Therefore, the next question is how to predict the SER at next time slot based on the channel estimates in current time slot. Back to our assumption, this is equivalent predicting the SER $\hat{P}_{s,n+2}$ at the $(n+2)$-th time slot, based on the channel estimation at $(n+1)$-th time slot (i.e. $\hat{\mathbf{h}}_{1,t,n+1}$, $\hat{\mathbf{h}}_{2,t,n+1}$).

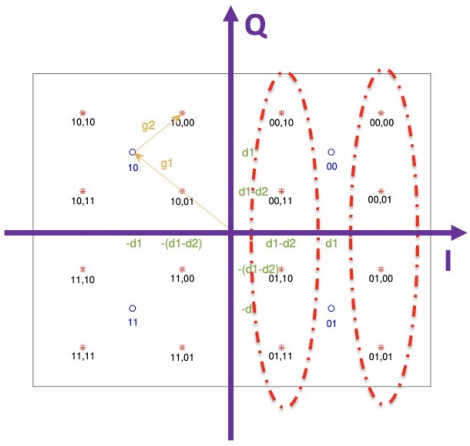

Figure 5.5: Overlapped constellation

Assume equally likely transmitted QPSK symbols for both users. Figure 5.5 shows that user 1's signal constellation is in presence of the other user's QPSK signal constellation as IUI. Here denote each constellation point is represented by four bits (i_1, i_2, i_3, i_4), where (i_1, i_2) is the QPSK symbol of user 1 and (i_3, i_4) denotes the one of user 2. In this case, as user 1 has stronger energy, it implies that user 1 is closer to BS, user 2 moves towards user 1 and becomes closer to the BS. Recall equation (5.9)

$$\tilde{\mathbf{x}}_{1,n+2} = \frac{\mathbf{Y}_{n+2}^T \hat{\mathbf{h}}_{1,t,n+1}^*}{\|\hat{\mathbf{h}}_{1,t,n+1}\|^2} = \frac{\mathbf{x}_{1,n+2}\mathbf{h}_{1,d,n+2}^T \hat{\mathbf{h}}_{1,t,n+1}^*}{\|\hat{\mathbf{h}}_{1,t,n+1}\|^2} + \frac{\mathbf{x}_{2,n+2}\mathbf{h}_{2,d,n+2}^T \hat{\mathbf{h}}_{1,t,n+1}^*}{\|\hat{\mathbf{h}}_{1,t,n+1}\|^2} + \mathbf{w}_1 = g_1\mathbf{x}_1 + g_2\mathbf{x}_2 + \mathbf{w}_1, \tag{5.25}$$

where $g_1 = \frac{\mathbf{h}_{1,d,n+2}^T \hat{\mathbf{h}}_{1,t,n+1}^*}{\|\hat{\mathbf{h}}_{1,t,n+1}\|^2}$ and $g_2 = \frac{\mathbf{h}_{2,d,n+2}^T \hat{\mathbf{h}}_{1,t,n+1}^*}{\|\hat{\mathbf{h}}_{1,t,n+1}\|^2}$. Therefore, we can see that $d_1 = \sqrt{g_1^2/2}$, $d_2 = \sqrt{g_2^2/2}$, and $\mathbf{w}_1$ is additive noise with each element w as CSCG random variables with variance σ^2. It is clear that g_1 represents the energy of user 1 and g_2 represents the normalized interference between user 1 and user 2 since we still apply MF detector here without taking IUI issue into account. Intuitively, ZF detector is generally applied to suppress

the IUI in multi-user scenario. To this end, denote the whole estimated channel matrix between the BS and both users at the $(n+1)$-th time slot as $\hat{\mathbf{H}}_{t,n+1} = \left[\hat{\mathbf{h}}_{1,t,n+1}^T, \hat{\mathbf{h}}_{2,t,n+1}^T\right]^T$, then the equation (5.9) for user 1 in the data detection phase is rewritten as

$$\tilde{\mathbf{x}}_{1,n+2} = \left(\hat{\mathbf{H}}_{t,n+1}^H \hat{\mathbf{H}}_{t,n+1}\right)_{(1,:)}^{-1} \hat{\mathbf{H}}_{t,n+1}^H \left(\mathbf{h}_{1,d,n+2}\mathbf{x}_{1,n+2}^T + \mathbf{h}_{2,d,n+2}\mathbf{x}_{2,n+2}^T\right) + \mathbf{w}_{1,n+2}, \tag{5.26}$$

where $\mathbf{w}_{1,n+2} = \left(\hat{\mathbf{H}}_{t,n+1}^H \hat{\mathbf{H}}_{t,n+1}\right)_{(1,:)}^{-1} \hat{\mathbf{H}}_{t,n+1}^H \mathbf{n}_{1,n+2}$ whose elements is still CSCG random variables with variance σ^2. In this ZF case, it is clear to rewrite the user 1's strength g_1 and the interference residue g_2 as

$$g_1 = \left(\hat{\mathbf{H}}_{t,n+1}^H \hat{\mathbf{H}}_{t,n+1}\right)_{(1,:)}^{-1} \hat{\mathbf{H}}_{t,n+1}^H \mathbf{h}_{1,d,n+2}, \quad g_2 = \left(\hat{\mathbf{H}}_{t,n+1}^H \hat{\mathbf{H}}_{t,n+1}\right)_{(1,:)}^{-1} \hat{\mathbf{H}}_{t,n+1}^H \mathbf{h}_{2,d,n+2}. \tag{5.27}$$

For both detectors, g_1 denotes the desired signal strength, g_2 denotes the interference (or interference residual) and $\mathbf{w}_1$ is the equivalent noise vector which follows the same distribution. As shown in Figure 5.5, for user 1, the blue dots are desired symbols. However, each blue dot could be one of the four red dots surrounding it due to the interference residue. Taking the additive CSCG noise into account, as long as the random noise does not take the constellation points across the I or Q axis, the symbols are able to be detected correctly. Based on this method, to obtain the closed-form expression of BER, we need to consider the worst case where the phase shift of the whole constellation of user 2 approaches $\frac{\pi}{2}$ as

shown in Figure . The BER P_{i_1} is given by

$$\begin{aligned} P_b = P_{i_1} = P_{i_2} &= \frac{1}{4}\left(P_r\left(d_1 - \sqrt{2}d_2 < |w|\right) + P_r\left(d_1 + \sqrt{2}d_2 < |w|\right)\right) \\ &= \frac{1}{2}\left(P_r\left(\frac{\sqrt{2}}{\sigma}w_R > \frac{\sqrt{2}}{\sigma}(d_1 - \sqrt{2}d_2)\right) + P_r\left(\frac{\sqrt{2}}{\sigma}w_R > \frac{\sqrt{2}}{\sigma}(d_1 + \sqrt{2}d_2)\right)\right) \\ &= \frac{1}{2}\left(Q\left(\frac{\sqrt{2}d_1 - 2d_2}{\sigma}\right) + Q\left(\frac{\sqrt{2}d_1 + 2d_2}{\sigma}\right)\right), \end{aligned} \tag{5.28}$$

where w_R represents the real part of the complex random variable w, and Q function is generally defined as $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-u^2/2)\,du$. Then it is straightforward to calculate SER P_s for QPSK modulation, based on the assumption $P_b^2 \ll P_b$

$$\begin{aligned} P_s &= 1 - (1 - P_b)(1 - P_b) \\ &= 2P_b - P_b^2 \\ &= Q\left(\frac{\sqrt{2}d_1 - 2d_2}{\sigma}\right) + Q\left(\frac{\sqrt{2}d_1 + 2d_2}{\sigma}\right). \end{aligned} \tag{5.29}$$

To predict the SER at the $(n+2)$-th time slot, we assume $\mathbf{h}_{1,d,n+2} \approx \hat{\mathbf{h}}_{1,t,n+1}$ based on the high spatial correlation between adjacent time slots. Therefore, associating parameters with the current time slot, we get

$$\hat{g}_1 = 1,\ \hat{g}_2 = \frac{\hat{\mathbf{h}}_{2,t}^T \hat{\mathbf{h}}_{1,t}^*}{|\hat{\mathbf{h}}_{1,t}|^2},\quad \hat{d}_1 = \sqrt{\frac{1}{2}},\quad \hat{d}_2 = \sqrt{\frac{\hat{g}_2^* \hat{g}_2}{2}}. \tag{5.30}$$

The resulting estimated SER for $(n+2)$-th time slot will be obtained based on the $(n+1)$-th time slot estimates

$$\hat{P}_s = Q\left(\frac{\sqrt{2}\hat{d}_1 - 2\hat{d}_2}{\sigma}\right) + Q\left(\frac{\sqrt{2}\hat{d}_1 + 2\hat{d}_2}{\sigma}\right). \tag{5.31}$$

All the above process can be adapted to user 2 directly. Recall the ignored subscripts, in the previous section, we already have estimate candidates $\hat{\mathbf{h}}_{1,t,n+1}$ and $\hat{\mathbf{h}}_{2,t,n+1}$ at the $(n+1)$-th time slot. In this *update* phase, we need to predict the SER in next time slot through equation (5.31). If the predicted SER does not exceed a threshold ε, which means the detected data stream is still trustable, $\hat{\mathbf{h}}_{1,t,n+1}$ and $\hat{\mathbf{h}}_{2,t,n+1}$ will be updated as $\hat{\mathbf{h}}_{1,t,n}$ and $\hat{\mathbf{h}}_{2,t,n}$. This whole tracking process iterates as long as the SER threshold is not satisfied.

The whole proposed data-aided channel tracking algorithm for multi-user scenario is summarized in Algorithm 3.

Algorithm 3 Multi-user Data-aided Channel Tracking Algorithm.

Input:

Truncated channel estimates of each user at the n-th time slot, $\hat{\mathbf{h}}_{k,t,n}$

Uplink received signal at the BS at the $(n+1)$-th time slot, $\mathbf{Y}_{n+1}$

Output:

Truncated channel estimates of each user at the $(n+1)$-th time slot, $\hat{\mathbf{h}}_{k,t,n+1}$; or scheduling method invoked and initial pilot-based channel estimation performed

1: Data detection phase to obtain $\tilde{\mathbf{x}}_{k,n+1}$ through (5.8) via MF detector or through (5.26) via ZF detector;
2: Channel decoding phase to obtain $\hat{\mathbf{x}}_{k,n+1}$;
3: Data-aided channel estimation phase to obtain $\hat{\mathbf{h}}_{k,d,n+1}$ through (5.10);
4: *Transition* phase to obtain the estimates of parameters $\hat{\alpha}_{n+1}$ and $\hat{\beta}_{n+1}$ through (5.18) via the Newtons iterative method;
5: Reconstruct the entire channel estimate at the $(n+1)$-th time slot $\hat{\mathbf{h}}_{k,n+1}$ through (5.22) based on the results from step 4;
6: Find the active antenna indices through Algorithm 2 and obtain the corresponding individual truncated channel estimate at the $(n+1)$-th time slot $\hat{\mathbf{h}}_{k,t,n+1}$;
7: Estimate the SER performance at the $(n+2)$-the time slot through (5.31), and check the threshold ϵ
8: **if** $\hat{P}_s \leq \epsilon$ **then**
9: *Update* phase processed
10: **else**
11: Scheduling method invoked and initial pilot-based channel estimation performed, then back to step 1
12: **end if**
13: **return** $\hat{\mathbf{h}}_{k,t,n+1}$;

5.4 Simulation results and outcome analysis

In this section, the numerical results for the proposed multi-user data-aided channel tracking system is present from different perspectives.

Assume two users in the same cell with the speeds of 150 km per hour and 360 km per hour respectively to simulate the overtaking scenario. The reason of choosing large speed

difference is to avoid handover operation between different cells, where user 1 is fast enough to to chase and pass user 2 before any of the two moves out of the range of the current cell. On the other hand, these speeds are typical for high-speed railway applications. The radius of the cell in mmWave communications is still 250 m, the number of the antennas on DLA at the BS is $M = 128$, and the number of active antenna for each user is $V = 3$, thus KV antennas in total at each time slot. Here we also assume that the symbol duration is 10^{-6} second and the typical block length is 1000, which means that for every time slot, the duration of the data block is 1 ms. As for mmWave communications with LoS path, the typical path loss exponent from the experiments is within the range of $[1.7, 2.7]$, where 2 is chosen for this simulation. We set a reasonable value of the SNR as 5 since none of the sophisticated channel coding methods is conducted in the simulation. The number of the simulated time slots is 4000 to prevent the vehicle from moving out of the cell. According to this setup, the trajectory of this overtaking scenario is shown in Figure 5.6, where user 1 overtakes user 2. The details of this co-channel issue analysis has been already discussed in Section 5.2.3, which is shown in Figure 5.7, following the sinc function shape. As we can see, the co-channel issue becomes severe, in terms of high channel correlation, around the 2000-th time slot, which means two users are getting closer and distinguishable during this time period.

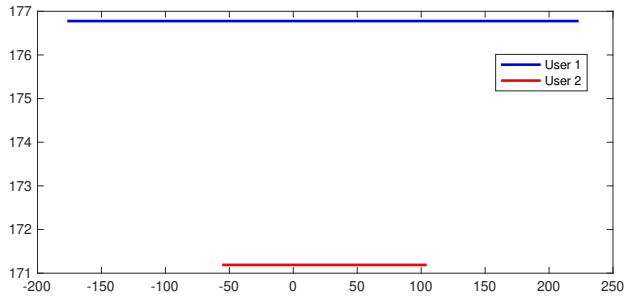

Figure 5.6: Trajectory setup for two-user scenario.

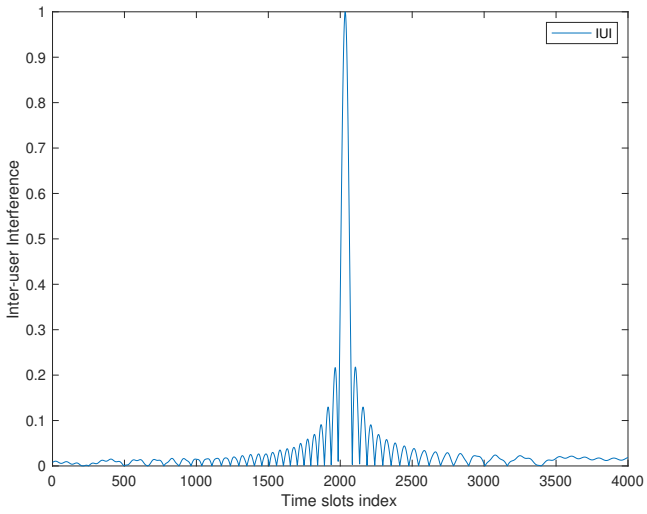

Figure 5.7: Co-channel issue evaluated by the channel correlation.

There are two important metrics to evaluate the overall performance of the system. One is the SER performance of data detection, and the other is the MSE (Mean Square Error) of the channel estimation, which are correlated with each other. The MSE of the channel tracking for both users is shown in Figure 5.9, while the SER performance for

both users is shown in Figure 5.8. At the beginning of the simulation, the two users are well separated with negligible interference, so noise dominates the channel tracking error. This is the equivalent situation with single-user scenario in the previous chapter. As the channel correlation becomes larger, the data detection result firstly gets worse. This portion of error is propagated and cause a mismatching via jointly finding the active antenna indices for the next time slot. As we can see from Figure 5.8, around 2000-th time slot when IUI (Inter-User Interference) dominates, the SER performance sharply degrades, thus causing large estimation error in Figure 5.9. Once the MSE of the channel estimation becomes untrustworthy, the error propagation issue, especially the mismatching between estimated and true active antenna indices, can not be solved until we perform another pilot-based channel training to extract the accurate channel estimation to restart the channel tracking system.

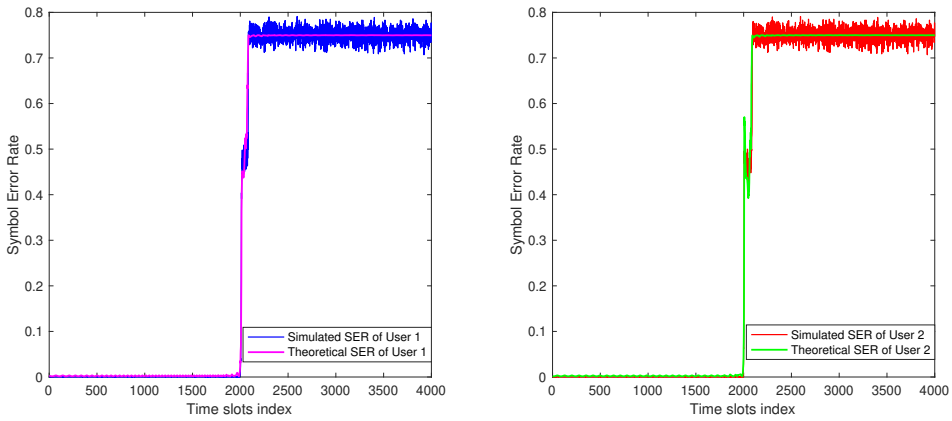

Figure 5.8: SER performance for both users

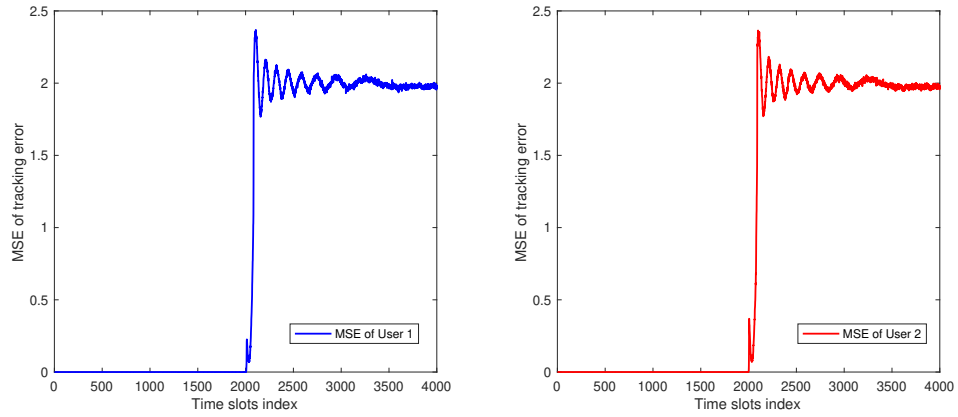

Figure 5.9: MSE of tracking error for both users

As the proposed Algorithm 3 shows, the *Update* phase is used to estimate the SER performance for the next time slot to equivalently see the channel correlation between both user. Firstly, to verify the proposed theoretical SER performance, Figure 5.8 also presents the theoretical SER at every time slot. As we can clearly see, the simulation result fluctuates but follows the theoretical result very well.

Next, we will show if the estimated SER is valid to be the scheduling metric, or equivalently to say how accurate by using estimated SER to approximate the true SER. This is related to the channel tracking error. If the MSE gets larger, it is not proper to use estimated channel vector to approximate the true channel vector in the next time slot, thus causing inaccurate SER estimation. Figure 5.10 illustrates this accuracy with different percentage of channel correlation under the proposed channel tracking systems with ZF detector. As shown in Figure 5.10, all the blue lines indicate the simulation results, red dots show the theoretical SER based on the true channel vector, while all the black lines indicate the estimated SER based on the truncated channel estimates of the $(n+1)$-th

time slot. Here, we compare the results in the cases based on different channel correlation percentage. As we can see, as the channel correlation goes higher, the simulation results always match the theoretical SER, which also verifies our derivation of the SER analysis in *Update* phase. As the co-channel issue becomes severe, the estimated SER suffers from larger offset from the true SER performance.

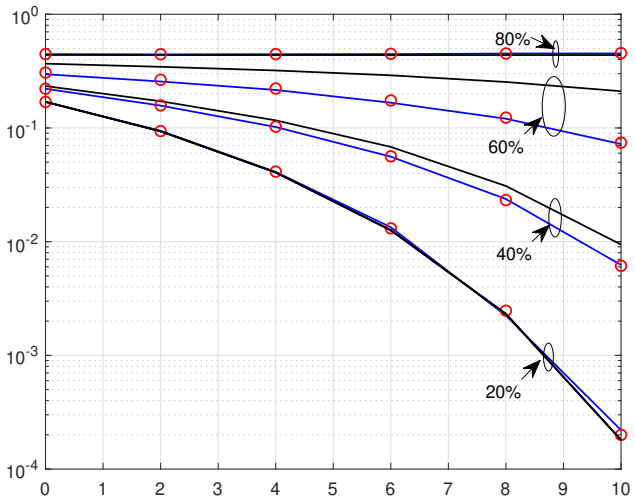

Figure 5.10: Estimated SER performance with different channel correlation.

5.5 Summary

In this chapter, we extend the previous work with multi-user scenario by considering inter-user interference. Firstly the spatial resolution problem is elaborated which is the greatest motivation for the research reported in this chapter. Secondly, we propose and advanced data-aided channel tracking system with scheduling decision based on SER prediction.

To achieve the SER prediction, the theoretical SER performance is derived for two-user scenario and has been verified by the simulation results. Lastly, the scheduling decision can be made based on the analysis of the estimated SER offset from the true SER performance. To thoroughly investigate channel tracking systems in different applications, we will present UPA-based data-aided channel tracking algorithm to explore better resolution in the 3D real life.

Chapter 6

UPA-based data-aided channel tracking framework

In the previous chapters, we have already proposed data-aided channel tracking algorithms for mmWave massive MIMO systems with ULA and DLA respectively. In the modern cellular network, ULA is the most popular antenna type and has been widely investigated in various of methods to acquire the CSI. Furthermore, DLA-based systems have attracted interest in recent years due to the low cost and complexity. However, they can only distinguish users along the horizontal direction. As the pico-cell becomes even smaller in 5G cellular network, the resolution in the vertical direction is necessary especially in urban area. For example, BS should be able to distinguish and schedule different users on the 3rd and 5th floor. Therefore, mmWave massive MIMO systems with UPA is being considered in this chapter.

6.1 UPA-based channel response

In this section, we consider a massive MIMO system with a UPA equipped at the BS and K single-antenna users in a single cell. At the BS, we assume a 3D coordinate, as shown in Figure 6.1, where the UPA is placed in the xz-plane and the reference antenna element $(0, 0)$ is placed at the origin. The number of antennas on the UPA is $M_{upa} = M_x \times M_z$, which are equally spaced on both vertical and horizontal directions with distance d. Instead of defining the conventional polar and azimuth angles, we define θ and ψ as the vertical angle and the angle between received signal and x-axis. To this end, the k-th user with arbitrary DoA can be represented as a coordinate of (d_k, θ_k, ψ_k), where d_k denotes the distance between the user and BS.

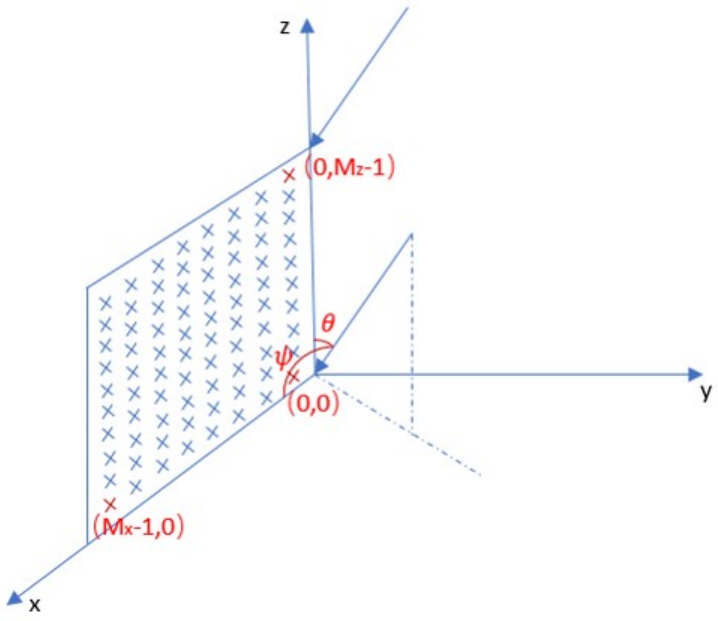

Figure 6.1: UPA configuration

As shown in Figure 6.1, we investigate the user with arbitrary DoA (θ, ψ). The propagation delay at the reference antenna element is defined as $(\tau_{x,0}, \tau_{z,0})$. Along x-axis, the delay of the antennas located at m_x-th column, where $m_x \in \{0, 1, ..., M_x - 1\}$, is denoted as $\tau_{x,m} = \tau(x, 0) - \frac{m_x d \cos\psi}{c}$. Therefore the horizontal steering vector is given by

$$\mathbf{v}_x = \frac{1}{\sqrt{M_x}} \left[1, e^{\frac{j2\pi m_x d \cos\psi}{\lambda}}, ..., e^{\frac{j2\pi(M_x-1) d \cos\psi}{\lambda}}\right]^T. \tag{6.1}$$

Similarly, along the z-axis, the delay of the antennas located at the m_z-th row, where $m_z \in \{0, 1, ..., M_z - 1\}$, is denoted as $\tau_{z,m} = \tau(z, 0) - \frac{m_z d \cos\theta}{c}$. Thus the vertical steering vector is given by

$$\mathbf{v}_z = \frac{1}{\sqrt{M_z}} \left[1, e^{\frac{j2\pi m_z d \cos\theta}{\lambda}}, ..., e^{\frac{j2\pi(M_z-1) d \cos\theta}{\lambda}}\right]^T. \tag{6.2}$$

As a result, the channel matrix between the user and BS $\mathbf{H} \in \mathbb{C}^{M_x \times M_z}$ under slow fading is given by

$$\mathbf{H} = \beta e^{j\phi} \mathbf{A} = \beta e^{j\phi} \left(\mathbf{v}_z \otimes \mathbf{v}_x^T\right), \tag{6.3}$$

where β is the attenuation scale factor, $e^{j\phi}$ denotes the random phase, and $\mathbf{A} \in \mathbb{C}^{M_x \times M_z}$ is the steering matrix with each element given as $\mathbf{A}(m_x, m_z) = e^{\frac{j2\pi d(m_x \cos\psi + m_z \cos\theta)}{\lambda}}$.

6.2 Data-aided channel tracking system

6.2.1 Initial channel estimation

At the 0-th time slot, we still apply the uplink pilot-based channel training to get the initial channel estimate, where TDD is assumed to explore the channel reciprocity. Suppose that

the initial channel training is conducted using orthogonal pilot sequence for the k-th user $\mathbf{p}_k \in \mathbb{C}^{Q\times 1}$, with length of Q and $q \in \{0, 1, ..., Q-1\}$. Then the uplink received signal at q-th symbol duration $\mathbf{Y}_{0,q} \in \mathbb{C}^{M_x \times M_z}$ is given by

$$\mathbf{Y}_{0,q} = \sum_{k=1}^{K} \mathbf{H}_{0,k} p_{k,q} + \mathbf{N}_{0,q}, \tag{6.4}$$

where $\mathbf{H}_{0,k}$ denotes the channel matrix at 0-th time slot, $p_{k,q}$ denotes the q-th symbol and $\mathbf{N}_{0,q}$ is CSCG noise with variance σ_n^2. By conducting least square estimator, the channel estimate of the k-th user is given as

$$\begin{aligned}\hat{\mathbf{H}}_{0,k} &= \frac{\sum_{q=0}^{Q-1} \mathbf{Y}_{0,q} p_{k,q}^*}{\sum_{q=0}^{Q-1} |p_{k,q}|^2} = \frac{\mathbf{H}_{0,k} \sum_{q=0}^{Q-1} p_{k,q} p_{k,q}^*}{\|\mathbf{p}_k\|^2} + \frac{\sum_{k'\neq k} \mathbf{H}_{0,k'} \sum_{q=0}^{Q-1} p_{k',q} p_{k,q}^*}{\|\mathbf{p}_k\|^2} + \frac{\sum_{q=0}^{Q-1} \mathbf{N}_{0,q} p_{k,q}^*}{\|\mathbf{p}_k\|^2} \\ &= \mathbf{H}_{0,k} + \mathbf{N}_{0,k},\end{aligned} \tag{6.5}$$

where $\mathbf{N}_{0,k} = \frac{\sum_{q=0}^{Q-1} \mathbf{N}_{0,q} p_{k,q}^*}{\|\mathbf{p}_k\|^2}$ is still CSCG noise. This initial channel estimate $\hat{\mathbf{H}}_{0,k}$ is further used for the following data detection and channel tracking process corresponding to the next time slot. In the following process, we assume single user scenario to explore tracking properties, thus getting rid of the subscript denoting user index to have $\hat{\mathbf{H}}_0$ for data detection in the next time slot.

6.2.2 Data-aided channel estimation

At the $(n+1)$-th time slot, suppose the uplink data stream $\mathbf{x}_{n+1} \in \mathbb{C}^{L\times 1}$ has a block length of L, and $l \in \{0, 1, ..., L-1\}$, then the received signal at l-th symbol duration $\mathbf{Y}_{n+1,l} \in \mathbb{C}^{M_x \times M_z}$

is given by

$$\mathbf{Y}_{n+1,l} = \mathbf{H}_{n+1} x_{n+1,l} + \mathbf{W}_{n+1,l}, \tag{6.6}$$

where $\mathbf{H}_{n+1}$ denotes the true channel matrix at $(n+1)$-th time slot, $x_{n+1,l}$ is the l-th symbol and $\mathbf{W}_{n+1,l}$ means CSCG noise with variance σ_n^2. To this end, we have the entire received data matrix during the $(n+1)$-th time slot $\mathbf{Y}_{n+1} \in \mathbb{C}^{M_x M_z \times L}$ with the help of vectorization on Equation (6.6)

$$\mathbf{Y}_{n+1} = \left[\text{vect}\left\{\mathbf{Y}_{n+1,0}\right\}, ..., \text{vect}\left\{\mathbf{Y}_{n+1,L-1}\right\}\right]. \tag{6.7}$$

Since the motion is normally linear and the channels are highly correlated between the adjacent time slots, we are able to recover the transmitted signal at $(n+1)$-th time slot by the assumption that $\hat{\mathbf{H}}_n \approx \mathbf{H}_{n+1}$. There is also a need to do the vectorization on $\hat{\mathbf{H}}_n$ to get $\hat{\mathbf{h}}_{n,v} \in \mathbb{C}^{M_x M_z \times 1}$ as

$$\hat{\mathbf{h}}_{n,v} = \text{vect}(\hat{\mathbf{H}}_n). \tag{6.8}$$

Then ZF detector is applied to recover the transmitted signal

$$\tilde{\mathbf{x}}_{n+1} = \left(\hat{\mathbf{h}}_{n,v}^H \hat{\mathbf{h}}_{n,v}\right)^{-1} \hat{\mathbf{h}}_{n,v}^H \mathbf{Y}_{n+1}. \tag{6.9}$$

After the sophisticated channel decoder, we will have the trustable data estimate $\hat{\mathbf{x}}_{n+1}$. After that, the data-aided channel estimation can be conducted

$$\hat{\mathbf{h}}_{n+1} = \frac{\mathbf{Y}_{n+1} \hat{\mathbf{x}}_{n+1}^*}{\|\hat{\mathbf{x}}_{n+1}\|^2}, \tag{6.10}$$

where $\hat{\mathbf{h}}_{n+1}$ is the vectorization of the channel matrix estimate at $(n+1)$-th time slot, which is defined as $\hat{\mathbf{H}}_{n+1} = \text{devect}\left(\hat{\mathbf{h}}_{n+1}\right)$.

6.2.3 *Transition* phase

In mmWave communications, where the LoS path is predominant, spatial sparsity is widely accepted to be exploited. Furthermore, with reasonable SNRs, the noise level could be generally stronger the NLoS paths due to extremely high attenuation and absorption. Therefore, the channel estimation issue is equivalent to estimate both the complex gain and spatial signature. Recall the channel model in Equation (6.3), we can rewrite it as

$$\mathbf{H}[m_x, m_z] = \beta_c e^{m_x\omega_x + m_z\omega_z}, \tag{6.11}$$

where β_c is the equivalent complex gain, $\omega_x = \frac{2\pi d \cos\psi}{\lambda}$ and $\omega_z = \frac{2\pi d \cos\theta}{\lambda}$. At this point, our target is to estimate four parameters as defined $\mathbf{\Psi} = [\beta_r, \beta_i, \omega_x, \omega_z]^T$, where β_r and β_i denote the real and imaginary part of the complex gain β respectively. Assume that the estimation error of the vectorization form of the channel matrix at $(n+1)$-th time slot, $\hat{\mathbf{h}}_{n+1}$, follows zero-mean Gaussian distribution, then we have the probability density function as follows

$$f(\hat{\mathbf{h}}_{n+1}|\beta_{n+1}, \omega_{x,n+1}, \omega_{z,n+1}) = (\pi\sigma^2)^{-M_x M_z} \exp\left\{-\frac{\|\hat{\mathbf{h}}_{n+1} - \mathbf{h}_{n+1}\|^2}{\sigma^2}\right\}, \tag{6.12}$$

where σ^2 is the variance of $\hat{\mathbf{h}}_{n+1}[u]$ with $u \in \{0, 1, ..., M_x M_z - 1\}$, following zero-mean Gaussian distribution as

$$E\left\{\left(\hat{\mathbf{h}}_{n+1} - \mathbf{h}_{n+1}\right)\left(\hat{\mathbf{h}}_{n+1} - \mathbf{h}_{n+1}\right)^H\right\} = \sigma^2 \mathbf{I}, \tag{6.13}$$

where $\mathbf{I}$ is a $M_x M_z \times M_x M_z$ identity matrix. As a result, the maximum likelihood estimate of $\mathbf{\Psi}$ is given as

$$\left[\hat{\beta}_{n+1}^{(ml)}, \hat{\omega}_{x,n+1}^{(ml)}, \hat{\omega}_{z,n+1}^{(ml)}\right] = \arg\max_{\mathbf{\Psi}} f(\hat{\mathbf{h}}_{n+1} | \beta_{n+1}, \omega_{x,n+1}, \omega_{z,n+1}) = \arg\min_{\mathbf{\Psi}} \|\hat{\mathbf{h}}_{n+1} - \mathbf{h}_{n+1}\|^2. \tag{6.14}$$

To obtain a close approximation to this optimal solution, we will apply a low complexity LS estimation via iterative Newton algorithm. To this end, we have the MLE of $\mathbf{\Psi}$ as

$$\hat{\mathbf{\Psi}}^{(ml)} = \arg\min_{\mathbf{\Psi}} \|\hat{\mathbf{h}}_r - \mathbf{h}_r\|^2 + \|\hat{\mathbf{h}}_i - \mathbf{h}_i\|^2 = \arg\min_{\mathbf{\Psi}} \|\hat{\mathbf{h}}_s - \mathbf{h}_s\|^2, \tag{6.15}$$

where $\hat{\mathbf{h}}_s$ and $\mathbf{h}_s$ are defined as

$$\hat{\mathbf{h}}_s = \left[\hat{\mathbf{h}}_r^T, \hat{\mathbf{h}}_i^T\right]^T, \mathbf{h}_s = \left[\mathbf{h}_r^T, \mathbf{h}_i^T\right]^T. \tag{6.16}$$

By using Newton method, we can get the estimate of $\mathbf{\Psi}$ in the $(j+1)$-th iteration, $\hat{\mathbf{\Psi}}^{(j+1)}$, based on the previous estimate in the j-th iteration $\hat{\mathbf{\Psi}}^{(j)}$ as

$$\hat{\mathbf{\Psi}}^{(j+1)} = \left(\mathbf{S}^H \mathbf{S}\right)^{-1} \mathbf{S}^H \left(\hat{\mathbf{h}}_s - \mathbf{s}\left(\hat{\mathbf{\Psi}}^{(j)}\right)\right) + \hat{\mathbf{\Psi}}^{(j)}, \tag{6.17}$$

where $\mathbf{s}(\mathbf{\Psi}) = \left[\mathbf{s}_r^T, \mathbf{s}_i^T\right]^T$ is given as

$$\mathbf{s}_r(\mathbf{\Psi})[u] = \beta_r e^{m_x\omega_x + m_z\omega_z}, \tag{6.18}$$

$$\mathbf{s}_i(\mathbf{\Psi})[u] = \beta_i e^{m_x\omega_x + m_z\omega_z}, \tag{6.19}$$

and $\mathbf{S} \in \mathbb{R}^{2M_xM_z \times 4}$ is the Jacobi matrix at $\mathbf{\Psi} = \hat{\mathbf{\Psi}}^{(j)}$:

$$\mathbf{S} = \left[\frac{\partial \mathbf{h}_r^T}{\partial \mathbf{\Psi}}, \frac{\partial \mathbf{h}_i^T}{\partial \mathbf{\Psi}}\right]^T \Bigg|_{\mathbf{\Psi}=\hat{\mathbf{\Psi}}^{(j)}} \tag{6.20}$$

This updated channel estimate is used for the data detection in the next time slot, and then the data-aided channel estimation and iterative tracking process are used to update the result based on current time slot. This track method is iterated to avoid resource-consuming pilot-based channel training.

To extract the initial estimate for the iterative method, we apply 2D DFT of the data-aided channel estimate $\hat{\mathbf{H}}_{n+1}$ as follows

$$\hat{\mathbf{H}}_\omega = \mathbf{F}_{M_z}\hat{\mathbf{H}}_{n+1}\mathbf{F}_{M_x}^T \approx \beta_c \mathbf{F}_{M_z}\mathbf{v}_z\mathbf{v}_x^T\mathbf{F}_{M_x}^T, \tag{6.21}$$

where the DFTs are defined as

$$\mathcal{X}_{\omega_x}[k] \triangleq \sum_{m_x=0}^{M_x-1} x[m_x]e^{-jm_xk\omega_{x,0}}, \tag{6.22}$$

$$\mathcal{X}_{\omega_z}[k] \triangleq \sum_{m_z=0}^{M_z-1} x[m_z]e^{-jm_zk\omega_{z,0}}, \tag{6.23}$$

where $\omega_{x,0} = \frac{2\pi}{M_x}$ and $\omega_{z,0} = \frac{2\pi}{M_z}$. Since there must be integer part and imaginary part

satisfying $\omega_x = (l_x + \alpha_x)\omega_{x,0}$ and $\omega_z = (l_z + \alpha_z)\omega_{z,0}$ respectively, where $l_x \in \{0, 1, ..., M_x - 1\}$, $l_z \in \{0, 1, ..., M_z - 1\}$, $\alpha_x \in [0, 1)$, and $\alpha_z \in [0, 1)$. The two strongest samples in each dimension, such as $\mathbf{v}_{\omega,x}[l_x]$ and $\mathbf{v}_{\omega,x}[l_x + 1]$, $\mathbf{v}_{\omega,z}[l_z]$ and $\mathbf{v}_{\omega,z}[l_z + 1]$, contain more than 80% of the total energy. Therefore, to roughly extract an initial estimate for the Newton iteration, we can simply choose the peak indices $m_{x,p}$ and $m_{z,p}$ to construct $\hat{\omega}_x = m_{x,p}\omega_{x,0}$ and $\hat{\omega}_z = m_{z,p}\omega_{z,0}$ respectively.

6.3 CRLB Analysis

In this section, we will analyze the CRLB of the channel estimation error in UPA scenario based on the Fisher matrix derivation. Here, to simplify the final CRLB expression, we redefine the targeting estimated parameters as $\mathbf{\Psi} = [\beta, \phi, \omega_x, \omega_z]$ by adopting the original channel model in Equation (6.3). Following the process described in the previous chapters, by getting rid of the time index subscripts, the probability density function of the vectorization form of the channel matrix $\hat{\mathbf{h}}$ is given by

$$f(\hat{\mathbf{h}}|\beta, \phi, \omega_x, \omega_z) = (\pi\sigma^2)^{-M_x M_z} \exp\left\{-\frac{\|\hat{\mathbf{h}} - \mathbf{h}\|^2}{\sigma^2}\right\}, \tag{6.24}$$

where σ^2 is the variance of $\hat{\mathbf{h}}[u]$ with $u \in \{0, 1, ..., M_x M_z - 1\}$, following zero-mean Gaussian distribution as

$$E\left\{\left(\hat{\mathbf{h}} - \mathbf{h}\right)\left(\hat{\mathbf{h}} - \mathbf{h}\right)^H\right\} = \sigma^2\mathbf{I}, \tag{6.25}$$

where $\mathbf{I}$ is a $M_x M_z \times M_x M_z$ identity matrix.

Then the log-likelihood function is given as

$$
\begin{aligned}
l(\hat{\mathbf{h}}|\beta,\phi,\omega_x,\omega_z) &= \ln f \\
&= -M_x M_z \ln\left(\pi\sigma^2\right) - \frac{\|\hat{\mathbf{h}}-\mathbf{h}\|^2}{\sigma^2} \\
&= -M_x M_z \ln\left(\pi\sigma^2\right) - \frac{\|\hat{\mathbf{h}}_r-\mathbf{h}_r\|^2}{\sigma^2} - \frac{\|\hat{\mathbf{h}}_i-\mathbf{h}_i\|^2}{\sigma^2},
\end{aligned}
\tag{6.26}
$$

which is based on the fact that $\|\hat{\mathbf{h}}-\mathbf{h}\|^2 = \|\hat{\mathbf{h}}_r-\mathbf{h}_r\|^2 + \|\hat{\mathbf{h}}_i-\mathbf{h}_i\|^2$.

According to the Fisher information matrix

$$
\mathbf{F} = E\left\{\begin{bmatrix}
\left(\frac{\partial l}{\partial\beta}\right)^2 & \frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\phi} & \frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\omega_x} & \frac{\partial l}{\partial\beta}\frac{\partial l}{\partial\omega_z} \\
\frac{\partial l}{\partial\phi}\frac{\partial l}{\partial\beta} & \left(\frac{\partial l}{\partial\phi}\right)^2 & \frac{\partial l}{\partial\phi}\frac{\partial l}{\partial\omega_x} & \frac{\partial l}{\partial\phi}\frac{\partial l}{\partial\omega_z} \\
\frac{\partial l}{\partial\omega_x}\frac{\partial l}{\partial\beta} & \frac{\partial l}{\partial\omega_x}\frac{\partial l}{\partial\phi} & \left(\frac{\partial l}{\partial\omega_x}\right)^2 & \frac{\partial l}{\partial\omega_x}\frac{\partial l}{\partial\omega_z} \\
\frac{\partial l}{\partial\omega_z}\frac{\partial l}{\partial\beta} & \frac{\partial l}{\partial\omega_z}\frac{\partial l}{\partial\phi} & \frac{\partial l}{\partial\omega_z}\frac{\partial l}{\partial\omega_x} & \left(\frac{\partial l}{\partial\omega_z}\right)^2
\end{bmatrix}\right\},
\tag{6.27}
$$

we need to derive the first-order partial derivatives as following

$$
\begin{aligned}
\frac{\partial l}{\partial\beta} &= \frac{2}{\sigma^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial\mathbf{h}_r}{\partial\beta} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial\mathbf{h}_i}{\partial\beta}\right], \\
\frac{\partial l}{\partial\phi} &= \frac{2}{\sigma^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial\mathbf{h}_r}{\partial\phi} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial\mathbf{h}_i}{\partial\phi}\right], \\
\frac{\partial l}{\partial\omega_x} &= \frac{2}{\sigma^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial\mathbf{h}_r}{\partial\omega_x} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial\mathbf{h}_i}{\partial\omega_x}\right], \\
\frac{\partial l}{\partial\omega_z} &= \frac{2}{\sigma^2}\left[(\hat{\mathbf{h}}_r-\mathbf{h}_r)^T\frac{\partial\mathbf{h}_r}{\partial\omega_z} + (\hat{\mathbf{h}}_i-\mathbf{h}_i)^T\frac{\partial\mathbf{h}_i}{\partial\omega_z}\right],
\end{aligned}
\tag{6.28}
$$

with the partial derivative of each antenna element with respect to the estimation param-

eters given by

$$\begin{aligned}
&\frac{\partial \mathbf{h}_r[u]}{\partial \beta} = \cos\left(m_x\omega_x + m_z\omega_z + \phi\right), \frac{\partial \mathbf{h}_i[u]}{\partial \beta} = \sin\left(m_x\omega_x + m_z\omega_z + \phi\right),\\
&\frac{\partial \mathbf{h}_r[u]}{\partial \phi} = -\beta\sin\left(m_x\omega_x + m_z\omega_z + \phi\right), \frac{\partial \mathbf{h}_i[u]}{\partial \phi} = \beta\cos\left(m_x\omega_x + m_z\omega_z + \phi\right),\\
&\frac{\partial \mathbf{h}_r[u]}{\partial \omega_x} = -\beta m_x\sin\left(m_x\omega_x + m_z\omega_z + \phi\right), \frac{\partial \mathbf{h}_i[u]}{\partial \omega_x} = \beta m_x\cos\left(m_x\omega_x + m_z\omega_z + \phi\right),\\
&\frac{\partial \mathbf{h}_r[u]}{\partial \omega_z} = -\beta m_z\sin\left(m_x\omega_x + m_z\omega_z + \phi\right), \frac{\partial \mathbf{h}_i[u]}{\partial \omega_z} = \beta m_z\cos\left(m_x\omega_x + m_z\omega_z + \phi\right).
\end{aligned} \tag{6.29}$$

Furthermore we also need to derive the expectations of the corresponding quadratic terms as shown

$$\begin{aligned}
E\left\{\left(\frac{\partial l}{\partial \beta}\right)^2\right\} &= \frac{2}{\sigma^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \beta}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \beta}\right|^2\right),\\
E\left\{\left(\frac{\partial l}{\partial \phi}\right)^2\right\} &= \frac{2}{\sigma^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \phi}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \phi}\right|^2\right)\\
E\left\{\left(\frac{\partial l}{\partial \omega_x}\right)^2\right\} &= \frac{2}{\sigma^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \omega_x}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \omega_x}\right|^2\right),\\
E\left\{\left(\frac{\partial l}{\partial \omega_z}\right)^2\right\} &= \frac{2}{\sigma^2}\left(\left|\frac{\partial \mathbf{h}_r}{\partial \omega_z}\right|^2 + \left|\frac{\partial \mathbf{h}_i}{\partial \omega_z}\right|^2\right),\\
E\left\{\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \phi}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \beta}\frac{\partial \mathbf{h}_r}{\partial \phi} + \frac{\partial \mathbf{h}_i^T}{\partial \beta}\frac{\partial \mathbf{h}_i^T}{\partial \phi}\right),\\
E\left\{\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \omega_x}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \beta}\frac{\partial \mathbf{h}_r}{\partial \omega_x} + \frac{\partial \mathbf{h}_i^T}{\partial \beta}\frac{\partial \mathbf{h}_i^T}{\partial \omega_x}\right),\\
E\left\{\frac{\partial l}{\partial \beta}\frac{\partial l}{\partial \omega_z}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \beta}\frac{\partial \mathbf{h}_r}{\partial \omega_z} + \frac{\partial \mathbf{h}_i^T}{\partial \beta}\frac{\partial \mathbf{h}_i^T}{\partial \omega_z}\right),\\
E\left\{\frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \omega_x}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \phi}\frac{\partial \mathbf{h}_r}{\partial \omega_x} + \frac{\partial \mathbf{h}_i^T}{\partial \phi}\frac{\partial \mathbf{h}_i^T}{\partial \omega_x}\right),\\
E\left\{\frac{\partial l}{\partial \phi}\frac{\partial l}{\partial \omega_z}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \phi}\frac{\partial \mathbf{h}_r}{\partial \omega_z} + \frac{\partial \mathbf{h}_i^T}{\partial \phi}\frac{\partial \mathbf{h}_i^T}{\partial \omega_z}\right),\\
E\left\{\frac{\partial l}{\partial \omega_x}\frac{\partial l}{\partial \omega_z}\right\} &= \frac{2}{\sigma^2}\left(\frac{\partial \mathbf{h}_r^T}{\partial \omega_x}\frac{\partial \mathbf{h}_r}{\partial \omega_z} + \frac{\partial \mathbf{h}_i^T}{\partial \omega_x}\frac{\partial \mathbf{h}_i^T}{\partial \omega_z}\right),
\end{aligned} \tag{6.30}$$

At this point, recall $\mathbf{\Psi} = [\beta, \phi, \omega_x, \omega_z]^T$, the Fisher information matrix can be rewritten as

$$\mathbf{F} = \frac{2}{\sigma^2}\mathbf{H}^T\mathbf{H}, \tag{6.31}$$

where the Jacobi matrix $\mathbf{H}$ is given as

$$\mathbf{H} = \left[\frac{\partial \mathbf{h}_r^T}{\partial \mathbf{\Psi}}, \frac{\partial \mathbf{h}_i^T}{\partial \mathbf{\Psi}}\right]^T. \tag{6.32}$$

To be specific, the rewritten Fisher information matrix is given as

$$\mathbf{F} = \frac{2}{\sigma^2}\begin{bmatrix} M_x M_z & 0 & 0 & 0 \\ 0 & \beta^2 M_x M_z & \beta^2 M_z \sum_{m_x} m_x & \beta^2 M_x \sum_{m_z} m_z \\ 0 & \beta^2 M_z \sum_{m_x} m_x & \beta^2 M_z \sum_{m_x} m_x^2 & \beta^2 \sum_{m_x,m_z} m_x m_z \\ 0 & \beta^2 M_x \sum_{m_z} m_z & \beta^2 \sum_{m_x,m_z} m_x m_z & \beta^2 M_x \sum_{m_z} m_z^2 \end{bmatrix}, \tag{6.33}$$

with the symmetrical elements

$$
\begin{aligned}
&\mathbf{F}[1,1] = \sum_{u} \cos^2\Delta + \sin^2\Delta = M_x M_z,\\
&\mathbf{F}[2,2] = \sum_{u} \beta^2\sin^2\Delta + \beta^2\cos^2\Delta = \beta^2 M_x M_z,\\
&\mathbf{F}[3,3] = M_z \sum_{m_x} \beta^2 m_x^2\sin^2\Delta + \beta^2 m_x^2\cos^2\Delta = \beta^2 M_z \sum_{m_x} m_x^2 \approx \beta^2 M_z \frac{M_x^3}{3},\\
&\mathbf{F}[4,4] = M_x \sum_{m_z} \beta^2 m_z^2\sin^2\Delta + \beta^2 m_z^2\cos^2\Delta = \beta^2 M_x \sum_{m_z} m_z^2 \approx \beta^2 M_x \frac{M_z^3}{3},\\
&\mathbf{F}[1,2] = \mathbf{F}[1,3] = \mathbf{F}[1,4] = 0,\\
&\mathbf{F}[2,3] = M_z \sum_{m_x} \beta^2 m_x\sin^2\Delta + \beta^2 m_x\cos^2\Delta = \beta^2 M_z \sum_{m_x} m_x = \beta^2 M_z \frac{M_x(M_x-1)}{2},\\
&\mathbf{F}[2,4] = M_x \sum_{m_z} \beta^2 m_z\sin^2\Delta + \beta^2 m_z\cos^2\Delta = \beta^2 M_x \sum_{m_z} m_z = \beta^2 M_x \frac{M_z(M_z-1)}{2},\\
&\mathbf{F}[3,4] = \sum_{u} \beta^2 m_x m_z\sin^2\Delta + \beta^2 m_x m_z\cos^2\Delta = \beta^2 \sum_{m_x,m_z} m_x m_z = \beta^2 M_z \frac{M_x M_z (M_x-1)(M_z-1)}{4},
\end{aligned}
\tag{6.34}
$$

where $\Delta = m_x\omega_x + m_z\omega_z + \phi$ and the approximation is based on the fact that $\sum_{n=0}^{N-1} n^2 = \frac{(N-1)N(2N-1)}{6} \approx \frac{N^3}{3}$ when N is large. To further explore the estimation error of each parameter in $\mathbf{\Psi}$, we need to derive the inverse of the Fisher information matrix based on the block matrix inversion theorem, which is given as

$$
\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}. \tag{6.35}
$$

Here we assume $\mathbf{A} = \begin{bmatrix} M_x M_z & 0 \\ 0 & \beta^2 M_x M_z \end{bmatrix}$, $\mathbf{B} = \mathbf{C}^T = \begin{bmatrix} 0 & 0 \\ \beta^2 M_z \sum_{m_x} m_x & \beta^2 M_x \sum_{m_z} m_z \end{bmatrix}$, and $\mathbf{D} = \beta^2 \begin{bmatrix} M_z \sum_{m_x} m_x^2 & \sum_{m_x,m_z} m_x m_z \\ \sum_{m_x,m_z} m_x m_z & M_x \sum_{m_z} m_z^2 \end{bmatrix}$. Then we are able to derive the diagonal elements of the inverse of the Fisher information matrix given as

$$\mathbf{F}^{-1}[1,1] = \frac{\sigma^2}{2\beta^2 M_x^2 M_z^2}\left(M_x M_z - \frac{3M_x(M_x-1)}{M_x+1} - \frac{3M_z(M_z-1)}{M_z+1}\right) \sim \frac{\sigma^2}{2\beta^2 M_x M_z}, \tag{6.36}$$

$$\mathbf{F}^{-1}[2,2] = \frac{\sigma^2}{2\beta^2 M_x M_z}, \tag{6.37}$$

$$\mathbf{F}^{-1}[3,3] = \frac{6\sigma^2}{\beta^2 M_z\left(M_x^3 - M_x\right)} \sim \frac{6\sigma^2}{\beta^2 M_z M_x^3}, \tag{6.38}$$

$$\mathbf{F}^{-1}[4,4] = \frac{6\sigma^2}{\beta^2 M_x\left(M_z^3 - M_z\right)} \sim \frac{6\sigma^2}{\beta^2 M_x M_z^3}. \tag{6.39}$$

As we can see that a larger UPA, along both vertical and horizontal directions, helps to increase the accuracy of the channel tracking. We can also see that the lower bounds of tracking error are all proportional to the variance of the noise σ^2, which also means to be inversely proportional to the length of the data sequence N. Therefore, it can improve the channel tracking accuracy by increasing the data block size.

6.3.1 Simulation results and outcome analysis

For the proposed data aided channel tracking system with UPA, there are some important parameters, the data block length N, the number of antennas at the BS/RSU (Road Side Unit) and SNR. Based on the above theoretical analysis, in this section, we will conduct simulations to evaluate the impacts of these parameters on channel tracking performance

with spatial signature estimation for both vertical and horizontal directions.

To set up the whole system, we still assume TDD to explore channel reciprocity, and single user scenario. Therefore, we only need to conduct uplink channel estimation, and downlink data transmission is able to apply the same estimate via feedback channel with negligible delay. Be consistent with the aforementioned setup environment, the BS/RSU is deployed with large UPA with $M_x \times M_z$ antenna elements, and only one user with single antenna is moving within the cell with speed of 360 km per hour, or equivalently 100 meter per second, which is the typical speed for the Chinese high-speed railway system. Throughout the simulation, the modulation scheme is QPSK, the block length of data transmission is N and the symbol duration is 1 ms. The length of the training pilot is set to be 100 to get decent initial channel estimation, and the path loss exponent is chosen to be 2 due to the LoS path only.

Next we will demonstrate the impacts on the accuracy performance associated with parameters M_x and M_z, in terms of the physical size of UPA array. As we can see from Figures 6.2, the numbers of antennas are varying with 25×25, 50×50 and 75×75. As we increase M, the accuracy can be enhanced by an order of magnitude. Recall the closed-form CRLBs of UPA-based channel tracking errors from equations (6.36) to (6.39), they are all inversely proportional to M_x and M_z, which is verified by our simulation results. Intuitively, more antennas at receiver side provide a higher power gain which improves the data detection accuracy, thus boosting the channel tracking performance. Especially for the accuracy of spatial signature estimation along both directions, the sparsity is more obvious as we increase the number of antennas in the corresponding direction as shown in Equations (6.38) and (6.39).

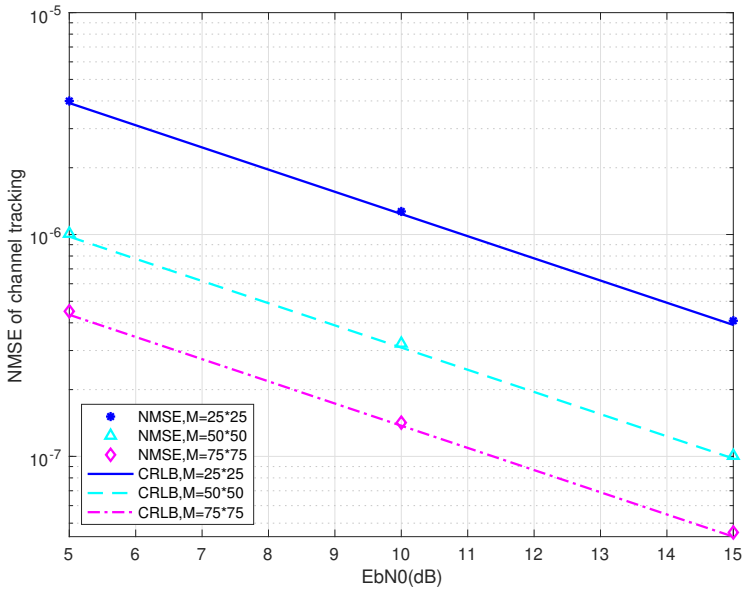

Figure 6.2: NMSE and CRLB of tracking error with different M.

Last but not the least, we will see the impact of the block length N on the channel tracking performance in Figure 6.3. Also recall the close-form CRLBs from equations (6.36) to (6.39), all of them are proportional to the noise variance σ^2, which means longer data sequence helps to suppress the noise. Every time we double the block length N, the NMSE increases by a factor of two. On the other hand, we are not able to use very large block length N since we need to use the previous channel estimates to approximate the current CSI to facilitate data-aided channel tracking process. As a result, the channel states of adjacent time slots have to be highly correlated, which requires a reasonable block length.

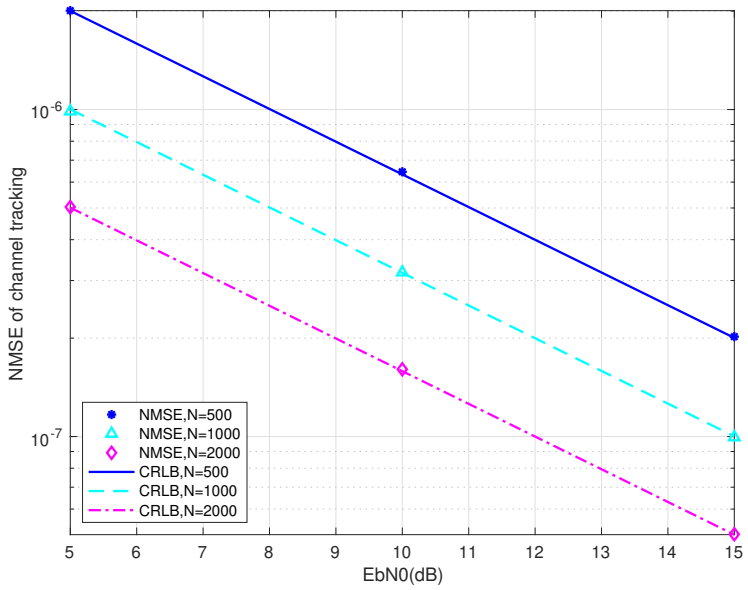

Figure 6.3: NMSE and CRLB of tracking error with different N.

6.4 Summary

In this chapter, a UPA-based data-aided channel tracking framework is presented for mmWave massive MIMO system with limited RF chains. Firstly the UPA array response has been carefully derived. Different from ULA case, 2D DFT is applied to extract spatial signatures on both vertical and horizontal directions. Based on this, Newton-method based channel tracking algorithm is also proposed, and the corresponding CRLB is derived with closed-form expressions. From the numerical results, we can clearly see that our proposed data-aided channel tracking system approaches the CRLB very well.

Chapter 7

Conclusions and future work

This thesis has presented the channel modeling, algorithm design and performance analysis on data-aided channel tracking for hybrid massive MIMO systems in millimeter wave communications. Thorough simulations have been conducted to verify the theoretical derivations.

7.1 Summary of contributions

- **ULA-based data-aided channel tracking**

  The ULA-based data-aided channel tracking method for mmWave massive MIMO system is presented. The basic idea of data-aided channel tracking is the high correlation between channels between the adjacent time slots. This makes it possible to employ the channel estimate from the previous time slot to detect the data stream in the current time slot. The theoretical close-form CRLBs with respect to spatial signature estimation and conventional DoA estimation are both derived, and verified

by the simulation results.

- **DLA-based data-aided channel tracking**

  The DLA-based data-aided channel tracking method for mmWave massive MIMO system is presented, which is a reproduction of a reference paper [1]. However, the closed-form CRLB is missing in this paper. Therefore, we carefully gave the detailed derivation of the closed-form expression for CRLB by solving the complicated summations with DFT properties. We also showed that DLA is not suitable for massive MIMO systems due to the overhead of the very large focal arc requirement. Even though it achieves dramatically low cost and complexity, the physical size of DLA is the biggest limitation for practical deployment.

- **DLA-based multi-user channel tracking**

  The DLA-based channel tracking method is extended for the multi-user case. The motivation of this work is the high channel correlation between users in the same beam where the BS is not able to distinguish them. To solve this problem, we need a metric to tell BS when to schedule user into different groups to avoid unacceptable channel tracking performance. The key idea is to estimate the SER performance in the next time slot via the channel estimate in the current time slot. The theoretical SER expression for two-user scenario is derived, and verified by the simulation results. Based on the estimated SER performance, simple group scheduling can avoid channel tracking degradation.

- **UPA-based data-aided channel tracking**

  A UPA-based data-aided channel tracking method for mmWave massive MIMO sys-

tem is proposed. The motivation of exploiting UPA antenna type is to adapt our channel tracking system to urban cellular network, where the BS need to distinguish different signal sources on the vertical direction. Followed by deriving the UPA array response, we found the 2D spatial sparsity and proposed 2D DFT to extract the spatial signatures on both vertical and horizontal directions. To reduce the complexity overhead, a truncated channel matrix is also employed. At last, the theoretical closed-form expression of CRLB is also derived and verified by the simulation results.

To summarize, the complexity comparison of three different antenna types is shown in Figure 7.1, where the complexities for inversion operation and FFT operation are presented with respect to entire channel representation and truncated channel representation respectively. The reason is that these two operations compose the majority of the overall complexity in data-aided channel tracking system we proposed.

	Inversion operation		**FFT operation**
	M	V	
ULA	$O(M^3)$	$O(V^3)$	$O(M \times \log(M))$
DLA	$O(M^3)$	$O(V^3)$	No complexity
UPA	$O(M_x^3 \times M_z^3)$	$O(V_x^3 \times V_z^3)$	$O(M_x \times M_z \times \log(M_x \times M_z))$

Figure 7.1: Complexity comparison of three different antenna types.

From Figure 7.1, we can see DLA-based system obviously is with the lowest consumption and complexity. However, as we have discussed in Chapter 4, the optical characteristics of DLA results in very large physical size by adapting massive MIMO, thus preventing it

from practical deployment. Therefore, we can conclude that ULA and UPA are still the commonly used antenna types in real life. To be specific, the combination of ULA and UPA is popular for urban area that benefits from the 2D resolution of UPA, whereas ULA only is a more economical choice for rural area.

7.2 Future work

In this thesis, the channel tracking systems with various antenna types have been investigated, and the multi-user scenario is also analyzed. As future work, the following issues could be considered.

- **Doppler compensation**

  Throughout the entire thesis, one of the critical assumptions is that Doppler effect has been compensated at the RF end by PLLs. However, this part would be quite challenging and requires complicated hardware design. A better solution is to compensate Doppler shift on the baseband signal. In the future, the joint spatial-signature-Doppler estimation is possible by considering the temporal sparsity as well in mmWave communications. The idea in [92] could be referred to, where 2D DFT is performed to do joint angle-Doppler estimation.

- **Sophisticated user scheduling**

  In this thesis, we proposed an estimated SER based scheduling metric, but did not propose any sophisticated user scheduling method. This related work could be done in the future by further taking achievable rate into account to guarantee the fairness and QoS. This is also an alive research area about exploiting beamforming, user scheduling

and resource allocation, especially for hybrid analog-digital communication systems with limited RF chains.

- **Downlink channel estimation**

  Throughout the thesis, another critical assumption is that only uplink channel tracking (estimation) is considered. For downlink data transmission, we assume a reliable feedback channel to feedback the estimated CSI to users without delay to perform data detection. In the future, the downlink channel estimation could also be investigated.

Bibliography

[1] Z. Gong, F. Jiang, and C. Li. Angle domain channel tracking with large antenna array for high mobility v2i millimeter wave communications. *IEEE Journal of Selected Topics in Signal Processing*, 13(5):1077–1089, 2019.

[2] Y. Zeng and R. Zhang. Millimeter wave mimo with lens antenna array: A new path division multiplexing paradigm. *IEEE Transactions on Communications*, 64(4):1557–1571, 2016.

[3] T. L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *Wireless Communications, IEEE Transactions on*, 9(11):3590–3600, Nov. 2010.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. Energy and spectral efficiency of very large multiuser mimo systems. *IEEE Transactions on Communications*, 61(4):1436–1449, 2013.

[5] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez. Millimeter wave mobile communications for 5G cellular: It will work! *IEEE Access*, 1:335–349, 2013.

[6] O. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath. Spatially sparse

precoding in millimeter wave mimo systems. *IEEE Transactions on Wireless Communications*, 13(3):1499–1513, 2014.

[7] W. Roh, J. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar. Millimeter-wave beamforming as an enabling technology for 5g cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine*, 52(2):106–113, 2014.

[8] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, NY, US, 2005.

[9] S. Han, C. I, Z. Xu, and C. Rowell. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5g. *IEEE Communications Magazine*, 53(1):186–194, 2015.

[10] J. Brady, N. Behdad, and A. M. Sayeed. Beamspace mimo for millimeter-wave communications: System architecture, modeling, analysis, and measurements. *IEEE Transactions on Antennas and Propagation*, 61(7):3814–3827, 2013.

[11] A. Goldsmith. *Wireless Communications*. Cambridge University Press, USA, 2005.

[12] L. Zheng and D.N.C. Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, 2003.

[13] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of mimo channels. *IEEE Journal on Selected Areas in Communications*, 21(5):684–702, 2003.

[14] J.H. Kotecha and A.M. Sayeed. Transmit signal design for optimal estimation of correlated mimo channels. *IEEE Transactions on Signal Processing*, 52(2):546–557, 2004.

[15] L. You, X. Gao, G. Y. Li, X. Xia, and N. Ma. BDMA for millimeter-wave/Terahertz massive MIMO transmission with per-beam synchronization. *IEEE Journal on Selected Areas in Communications*, 35(7):1550–1563, July 2017.

[16] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin. Angle domain hybrid precoding and channel tracking for millimeter wave massive mimo systems. *IEEE Transactions on Wireless Communications*, 16(10):6868–6880, 2017.

[17] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh. Millimeter wave beamforming for wireless backhaul and access in small cell networks. *IEEE Transactions on Communications*, 61(10):4391–4403, Oct. 2013.

[18] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic. Millimeter wave MIMO channel estimation using overlapped beam patterns and rate adaptation. *IEEE Transactions on Signal Processing*, 65(3):601–616, Feb. 2017.

[19] J. Singh and S. Ramakrishna. On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays. *IEEE Transactions on Wireless Communications*, 14(5):2670–2683, 2015.

[20] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin. Angle domain hybrid precoding and channel tracking for millimeter wave massive mimo systems. *IEEE Transactions on Wireless Communications*, 16(10):6868–6880, Oct 2017.

[21] A. Liu and V. Lau. Phase only rf precoding for massive mimo systems with limited rf chains. *IEEE Transactions on Signal Processing*, 62(17):4505–4515, 2014.

[22] L. Yang, Y. Zeng, and R. Zhang. Channel estimation for millimeter-wave MIMO communications with lens antenna arrays. *IEEE Transactions on Vehicular Technology*, 67(4):3239–3251, April 2018.

[23] A. Liu, V. Lau, and M. Honig. Compressive rf training for massive mimo with channel support side information. *IEEE Transactions on Wireless Communications*, 18(7):3628–3641, 2019.

[24] Z. Gong, C. Li, and F. Jiang. Channel estimation for sparse massive mimo channels in low snr regime. *IEEE Transactions on Cognitive Communications and Networking*, 4(4):883–893, 2018.

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

[26] S.K. Sahoo and A. Makur. Signal recovery from random measurements via extended orthogonal matching pursuit. *IEEE Transactions on Signal Processing*, 63(10):2572–2581, 2015.

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

[28] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28:253–263, 12 2008.

[29] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[30] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. 20(1), 1998.

[31] IU. E. Nesterov. *Interior-point polynomial algorithms in convex programming*. SIAM

studies in applied mathematics ; vol. 13. Society for Industrial and Applied Mathematics SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104, Philadelphia, Pa, 1994.

[32] W. Li and J. Preisig. Estimation of rapidly time-varying sparse channels. *IEEE Journal of Oceanic Engineering*, 32(4):927–939, 2007.

[33] D. L. Donoho, Y. Tsaig, I. Drori, and J. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58(2):1094–1121, 2012.

[34] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):310–316, 2010.

[35] H. Huang and A. Makur. Backtracking-based matching pursuit method for sparse signal reconstruction. *IEEE Signal Processing Letters*, 18(7):391–394, 2011.

[36] J. Wang, S. Kwon, P. Li, and B. Shim. Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis. *IEEE Transactions on Signal Processing*, 64(4):1076–1089, 2016.

[37] J. Lee, G. Gil, and Y. Lee. Channel estimation via orthogonal matching pursuit for hybrid mimo systems in millimeter wave communications. *IEEE Transactions on Communications*, 64(6):2370–2386, 2016.

[38] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath. Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave mimo systems. *IEEE Transactions on Wireless Communications*, 17(5):2946–2960, May 2018.

[39] M. N. Kulkarni, A. Ghosh, and J. G. Andrews. A comparison of mimo techniques in downlink millimeter wave cellular networks with hybrid beamforming. *IEEE Transactions on Communications*, 64(5):1952–1967, 2016.

[40] A. Morsali, A. Haghighat, and B. Champagne. Generalized framework for hybrid analog/digital signal processing in massive and ultra-massive-mimo systems. *IEEE Access*, 8:100262–100279, 2020.

[41] Z. Guo, X. Wang, and W. Heng. Millimeter-wave channel estimation based on 2-D beamspace MUSIC method. *IEEE Transactions on Wireless Communications*, 16(8):5384–5394, Aug. 2017.

[42] H. Lin, F. Gao, S. Jin, and G. Y. Li. A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access. *IEEE Journal on Selected Areas in Communications*, 35(10):2268–2280, Oct. 2017.

[43] D. Fan, F. Gao, Y. Liu, Y. Deng, G. Wang, Z. Zhong, and A. Nallanathan. Angle domain channel estimation in hybrid millimeter wave massive mimo systems. *IEEE Transactions on Wireless Communications*, 17(12):8165–8179, 2018.

[44] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan. Channel estimation and transmission strategy for hybrid mmwave noma systems. *IEEE Journal of Selected Topics in Signal Processing*, 13(3):584–596, 2019.

[45] J.B. Andersen, J. Jensen, S.H. Jensen, and F. Frederiksen. Prediction of future fading based on past measurements. In *Gateway to 21st Century Communications Village. VTC 1999-Fall. IEEE VTS 50th Vehicular Technology Conference (Cat. No.99CH36324)*, volume 1, pages 151–155 vol.1, 1999.

[46] J. K. Hwang and J.H. Winters. Sinusoidal modeling and prediction of fast fading

processes. In *IEEE GLOBECOM 1998 (Cat. NO. 98CH36250)*, volume 2, pages 892–897 vol.2, 1998.

[47] L. Dong, G. Xu, and H. Ling. Prediction of fast fading mobile radio channels in wideband communication systems. In *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No.01CH37270)*, volume 6, pages 3287–3291 vol.6, 2001.

[48] R. Vaughan, P. Teal, and R. Raich. Short-term mobile channel prediction using discrete scatterer propagation model and subspace signal processing algorithms. In *Vehicular Technology Conference Fall 2000. IEEE VTS Fall VTC2000. 52nd Vehicular Technology Conference (Cat. No.00CH37152)*, volume 2, pages 751–758 vol.2, 2000.

[49] J. Maurer, T. Fugen, and W. Wiesbeck. Narrow-band measurement and analysis of the inter-vehicle transmission channel at 5.2 ghz. In *Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367)*, volume 3, pages 1274–1278 vol.3, 2002.

[50] R. Roy and T. Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989.

[51] R. O. Adeogun, P. D. Teal, and P. A. Dmochowski. Extrapolation of mimo mobile-to-mobile wireless channels using parametric-model-based prediction. *IEEE Transactions on Vehicular Technology*, 64(10):4487–4498, 2015.

[52] M. Sharif and B. Hassibi. On the capacity of mimo broadcast channels with partial side information. *IEEE Transactions on Information Theory*, 51(2):506–522, 2005.

[53] M. Sharif and B. Hassibi. A comparison of time-sharing, dpc, and beamforming for

mimo broadcast channels with many users. *IEEE Transactions on Communications*, 55(1):11–15, 2007.

[54] D. Aktas, M.N. Bacha, J.S. Evans, and S.V. Hanly. Scaling results on the sum capacity of cellular networks with mimo links. *IEEE Transactions on Information Theory*, 52(7):3264–3274, 2006.

[55] T. L. Marzetta. How much training is required for multiuser mimo? In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pages 359–363, 2006.

[56] J. Vanderpypen and L. Schumacher. Mimo channel prediction using esprit based techniques. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5, 2007.

[57] H. Shirani-Mehr, D. N. Liu, and G. Caire. Channel state prediction, feedback and scheduling for a multiuser mimo-ofdm downlink. In *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pages 136–140, 2008.

[58] Z. Shen, K. Xu, Y. Wang, and W. Xie. Angle-domain channel tracking for high speed railway communications with massive ula. In *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, pages 159–165, 2018.

[59] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang. Fast channel tracking for terahertz beamspace massive mimo systems. *IEEE Transactions on Vehicular Technology*, 66(7):5689–5696, 2017.

[60] S.M. Kay and S.L. Marple. Spectrum analysis—a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981.

[61] A. Duel-Hallen, Hu. S., and H. Hallen. Long-range prediction of fading signals. *IEEE Signal Processing Magazine*, 17(3):62–75, 2000.

[62] K.E. Baddour and N.C. Beaulieu. Autoregressive modeling for fading channel simulation. *IEEE Transactions on Wireless Communications*, 4(4):1650–1662, 2005.

[63] P. Stoica, R. L. Moses, et al. Spectral analysis of signals. 2005.

[64] A. Duel-Hallen. Fading channel prediction for mobile radio adaptive transmission systems. *Proceedings of the IEEE*, 95(12):2299–2313, 2007.

[65] J. G. Proakis. *Digital Communications 5th Edition*. McGraw Hill, 2007.

[66] D. Schafhuber and G. Matz. Mmse and adaptive prediction of time-varying channels for ofdm systems. *IEEE Transactions on Wireless Communications*, 4(2):593–602, 2005.

[67] S. Semmelrodt and R. Kattenbach. Investigation of different fading forecast schemes for flat fading radio channels. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, volume 1, pages 149–153 Vol.1, 2003.

[68] C.Y. Fung and S.C. Chan. Estimation of fast fading channel in impulse noise environment. In *2002 IEEE International Symposium on Circuits and Systems. Proceedings (Cat. No.02CH37353)*, volume 4, pages IV–IV, 2002.

[69] K. E. Baddour, C. C. Squires, and T. J. Willink. Mobile channel prediction with application to transmitter antenna selection for alamouti systems. In *IEEE Vehicular Technology Conference*, pages 1–6, 2006.

[70] Z. Liu, X. Ma, and G.B. Giannakis. Space-time coding and kalman filtering for time-selective fading channels. *IEEE Transactions on Communications*, 50(2):183–186, 2002.

[71] B. Balakumar, S. Shahbazpanahi, and T. Kirubarajan. Joint mimo channel tracking and symbol decoding using kalman filtering. *IEEE Transactions on Signal Processing*,

55(12):5873–5879, 2007.

[72] H. Li, Y. Wang, M. Jiang, and D. Yuan. Doubly selective fading channel tracking based on particle filter in mimo-ofdm systems. In *2007 International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, pages 1084–1087, 2007.

[73] X. Zhang, P. Xiao, D. Ma, and J. Wei. Variational-bayes-assisted joint signal detection, noise covariance estimation, and channel tracking in mimo-ofdm systems. *IEEE Transactions on Vehicular Technology*, 63(9):4436–4449, 2014.

[74] Y. Wang, Y. Wang, S. Zhang, and H. Cen. Channel tracking and transmission design in 5g large-scale mimo system. *IEEE Access*, 7:62032–62041, 2019.

[75] R. Chopra and C. R. Murthy. Data aided mse-optimal time varying channel tracking in massive mimo systems. *IEEE Transactions on Signal Processing*, 69:4219–4233, 2021.

[76] J. K. Tugnait, S. He, and H. Kim. Doubly selective channel estimation using exponential basis models and subblock tracking. *IEEE Transactions on Signal Processing*, 58(3):1275–1289, 2010.

[77] C. Zhang, J. Zhang, Y. Huang, and L. Yang. Location-aided channel tracking and downlink transmission for hst massive mimo systems. *IET Commun.*, 11:2082–2088, 2017.

[78] K. Xu, Z. Shen, Y. Wang, and X. Xia. Location-aided mmimo channel tracking and hybrid beamforming for high-speed railway communications: An angle-domain approach. *IEEE Systems Journal*, 14(1):93–104, 2020.

[79] P. Kela, M. Costa, J. Turkka, M. Koivisto, J. Werner, A. Hakkarainen, M. Valkama,

R. Jantti, and K. Leppanen. Location based beamforming in 5g ultra-dense networks. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–7, 2016.

[80] J. Zhao, H. Xie, F. Gao, W. Jia, S. Jin, and H. Lin. Time varying channel tracking with spatial and temporal bem for massive mimo systems. *IEEE Transactions on Wireless Communications*, 17(8):5653–5666, 2018.

[81] I. M. Baby, K. Appaiah, and R. Chopra. Optimal channel tracking and power allocation for time varying fdd massive mimo systems. *IEEE Transactions on Communications*, 70(2):1229–1244, 2022.

[82] G.B. Giannakis and C. Tepedelenlioglu. Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proceedings of the IEEE*, 86(10):1969–1986, 1998.

[83] S. He and J. K. Tugnait. Decision-directed tracking of doubly-selective channels using exponential basis models. In *2008 IEEE International Conference on Communications*, pages 5098–5102, 2008.

[84] D.K. Borah and B.T. Hart. Frequency-selective fading channel estimation with a polynomial time-varying channel model. *IEEE Transactions on Communications*, 47(6):862–873, 1999.

[85] D. Hu, X. Wang, and L. He. A new sparse channel estimation and tracking method for time-varying ofdm systems. *IEEE Transactions on Vehicular Technology*, 62(9):4648–4653, 2013.

[86] M. Martone. Wavelet-based separating kernels for array processing of cellular ds/cdma signals in fast fading. *IEEE Transactions on Communications*, 48(6):979–995, 2000.

[87] T. Zemen and C.F. Mecklenbrauker. Time-variant channel estimation using discrete prolate spheroidal sequences. *IEEE Transactions on Signal Processing*, 53(9):3597–3607, 2005.

[88] F. Talaei, J. Zhan, and X. Dong. Low complexity mimo channel prediction for fast time-variant vehicular communications channels based on discrete prolate spheroidal sequences. *IEEE Access*, 9:23398–23408, 2021.

[89] A.M. Sayeed, A. Sendonaris, and B. Aazhang. Multiuser detection in fast-fading multipath environments. *IEEE Journal on Selected Areas in Communications*, 16(9):1691–1701, 1998.

[90] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.

[91] E. Fishler, M. Grosmann, and H. Messer. Detection of signals by information theoretic criteria: general asymptotic performance analysis. *IEEE Transactions on Signal Processing*, 50(5):1027–1036, 2002.

[92] Zijun Gong, Cheng Li, Fan Jiang, and Moe Z. Win. Data-aided doppler compensation for high-speed railway communications over mmwave bands. *IEEE Transactions on Wireless Communications*, 20(1):520–534, 2021.

[93] M. A. Al-Joumayly and N. Behdad. Wideband planar microwave lenses using sub-wavelength spatial phase shifters. *IEEE Transactions on Antennas and Propagation*, 59(12):4542–4552, 2011.

Appendix A

DLA-based array response

According to [2], the DLA-based array response follows 'sinc' function which explores the spatial sparsity. Next, we will follow the similar methodology in [2] to derive the array response mathematically. As shown in Figure 1.1, we denote ϕ_p as the physical angle of the uniform plane wave, $x_0(\phi_p)$ as the impinging signal at the center of the lens aperture, $r_m(\phi_p)$ as the resulting signal received by the m-th antenna on the focal arc, and $\mathbf{a}(\phi_p) \in \mathbb{C}^{M\times 1}$ as the array response whose elements are defined by the ratio $a_m(\phi_p) \triangleq r_m(\phi_p)/x_0(\phi_p)$. The methodology to derive the array response follows the properties of the EM lens. Let's set B_0 as the reference point, and denote the phase shift profile as $\Phi(y, x)$ to represent the phase delay relative to B_0 by the optical characteristics of the lens at any point $(0, y, z)$ on the aperture. Therefore, the goal is to achieve constructive superposition of the corresponding phase of all rays at reference point B_0 [93]. With $B_0(F, 0, 0)$ and any $x(0, y, z)$ on the lens, the distance between them can be calculated as $d(y, z, B_0)$, then we

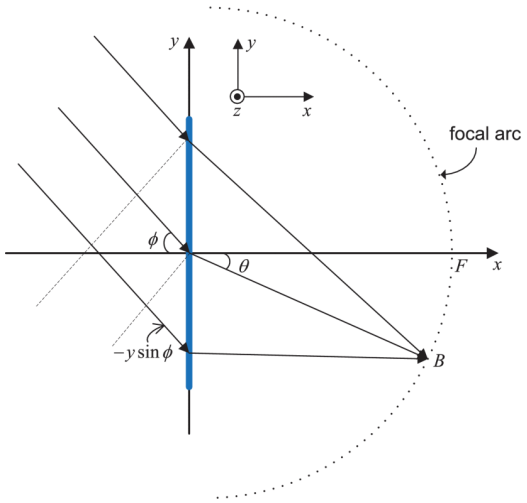

Figure 1.1: DLA response derivation [2].

have

$$\Phi(y,z) = \Phi_0 - k_0 d(y,z,B_0) = \Phi_0 - k_0\sqrt{F^2+y^2+z^2}, \tag{A.1}$$

$$k_0 = 2\pi/\lambda, \tag{A.2}$$

$$\forall (y,z) \in \left[-\frac{D_y}{2},\frac{D_y}{2}\right] \times \left[-\frac{D_z}{2},\frac{D_z}{2}\right]. \tag{A.3}$$

Based on the phase shift profile on the reference point B_0, for an arbitrary point $B(F\cos\theta, -F\sin\theta, 0)$ on the focal arc as shown in Figure 1.1, the resulting phase delay $\psi(y,z,B)$ is given by:

$$\begin{aligned}
\psi(y,z,B) &= \Phi(y,z) + k_0 d(y,z,B) \\
&= \Phi(y,z) + k_0\sqrt{(F\cos\theta)^2 + (-F\sin\theta - y)^2 + z^2} \\
&= \Phi(y,z) + k_0\sqrt{F^2(\cos\theta^2 + \sin\theta^2) + y^2 + z^2 + 2yF\sin theta} \\
&= \Phi_0 - k_0\sqrt{F^2+y^2+z^2} + k_0\sqrt{F^2(\cos\theta^2+\sin\theta^2)+y^2+z^2+2yF\sin\theta}.
\end{aligned} \tag{A.4}$$

If we take Taylor expansion on the third term, the original function of F is given by

$$f(F) = \sqrt{(F + y\sin\theta)^2 + y^2 + z^2 - y^2\sin^2\theta}. \tag{A.5}$$

Let $F = F + y\sin\theta$, then

$$f(F) = \sqrt{F^2 + y^2 + z^2 - y^2\sin^2\theta}, \tag{A.6}$$

which will be expanded on the point $F_0 = F$ with the first-order derivative of the original function is given by

$$\begin{aligned} \frac{\partial f(F)}{\partial F} &= \frac{1}{2}\frac{1}{\sqrt{F^2 + y^2 + z^2 - y^2\sin^2\theta}}2F \\ &= \frac{F}{\sqrt{F^2 + y^2 + z^2 - y^2\sin^2\theta}}. \end{aligned} \tag{A.7}$$

Based on the assumption $F \gg D_y, D_z$, higher-order terms are negligible, so it can be approximated as the first-order Taylor expansion as

$$\begin{aligned} f(F) &\approx f(F) + \frac{\partial f(F)}{\partial F}(F - F_0) \\ &= \sqrt{F^2 + y^2 + z^2 - y^2\sin^2\theta} + \frac{F}{\sqrt{F^2 + y^2 + z^2 - y^2\sin^2\theta}}y\sin\theta \\ &\approx \sqrt{F^2 + y^2 + z^2} + \frac{F}{\sqrt{F^2 + y^2 + z^2}}y\sin\theta^1 \\ &\approx \sqrt{F^2 + y^2 + z^2} + y\sin\theta^2. \end{aligned} \tag{A.8}$$

To this end, the resulting phase delay on point B of (A.4) can be rewritten as

$$\begin{aligned} \psi(y, z, B) &\approx \Phi_0 - k_0\sqrt{F^2 + y^2 + z^2} + k_0\sqrt{F^2 + y^2 + z^2} + k_0 y\sin\theta \\ &= \Phi_0 + k_0 y\sin\theta. \end{aligned} \tag{A.9}$$

Next, we will take this phase delay into account to see the resulting received signal. Let $s(y,z)$ denote the input signal on the EM lens coming from the direction of ϕ_p. For uniform incident plane waves, it is widely accepted that $s(y,z) = s(y)$ as elevation AoA is ignored. Recall that $x_0(\phi_p)$ is the impinging signal at the center of the lens aperture, then we have

$$s(y) = \frac{1}{\lambda\sqrt{D_y D_z}} x_0(\phi_p) e^{j\frac{2\pi}{\lambda} y \sin\phi_p}. \tag{A.10}$$

Then the received signal on the focal arc will be

$$\begin{aligned} r_m(\phi_p) &= \int_{-D_z/2}^{D_z/2} \int_{-D_y/2}^{D_y/2} s(y,z) e^{-j\psi(y,z,\theta_m)} dy dz \\ &\approx D_z \int_{-D_y/2}^{D_y/2} s(y) e^{-j(\Phi_0 + \frac{2\pi}{\lambda} y \sin\theta_m)} dy \\ &= D_z e^{-j\Phi_0} \int_{-D_y/2}^{D_y/2} s(y) e^{-j\frac{2\pi}{\lambda} y \sin\theta_m} dy. \end{aligned} \tag{A.11}$$

To facilitate our derivation, we denote $D = D_y/\lambda$, $\tilde{y} = y/\lambda$, $\tilde{\theta}_m = \sin\theta_m$, $\tilde{\phi}_p = \sin\phi_p$, and the linear scaling of the $s(y)$ as $\tilde{s}(\tilde{y}) = \lambda s(\lambda\tilde{y})$. Then the equation (A.11) will be rewritten as

$$\begin{aligned} r_m(\phi_p) &= D_z e^{-j\Phi_0} \int_{-D/2}^{D/2} \lambda \frac{1}{\lambda\sqrt{D_y D_z}} x_0(\phi_p) e^{j\frac{2\pi}{\lambda}\lambda\tilde{y}\tilde{\phi}_p} e^{-j2\pi\tilde{y}\tilde{\theta}_m} d\tilde{y} \\ &= \sqrt{\frac{D_z}{D_y}} x_0(\phi_p) e^{-j\Phi_0} \int_{-D/2}^{D/2} e^{-j2\pi\tilde{y}(\tilde{\theta}_m - \tilde{\phi}_p)} d\tilde{y} \\ &= \sqrt{\frac{D_z}{D_y}} x_0(\phi_p) e^{-j\Phi_0} \int_{-D/2}^{D/2} \cos\left(2\pi(\tilde{\theta}_m - \tilde{\phi}_p)\tilde{y}\right) - i\sin\left(2\pi(\tilde{\theta}_m - \tilde{\phi}_p)\tilde{y}\right) d\tilde{y} \\ &= \sqrt{\frac{D_z}{D_y}} x_0(\phi_p) e^{-j\Phi_0} \left[\frac{1}{2\pi(\tilde{\theta}_m - \tilde{\phi}_p)} \sin\left(2\pi(\tilde{\theta}_m - \tilde{\phi}_p)\tilde{y}\right) + \frac{i}{2\pi(\tilde{\theta}_m - \tilde{\phi}_p)} \cos\left(2\pi(\tilde{\theta}_m - \tilde{\phi}_p)\tilde{y}\right) \right]\Bigg|_{-D/2}^{D/2} \\ &= \sqrt{\frac{D_z}{D_y}} x_0(\phi_p) e^{-j\Phi_0} D \frac{\sin\left(\pi D(\tilde{\theta}_m - \tilde{\phi}_p)\right)}{\pi D(\tilde{\theta}_m - \tilde{\phi}_p)}. \end{aligned} \tag{A.12}$$

With $A \triangleq \frac{D_y D_z}{\lambda^2}$, $\tilde{\theta}_m \triangleq \sin\theta_m = m/D$ as $\tilde{\theta}_m$ is equally spaced in the interval $[-1, 1]$, and recall $D = D_y/\lambda$, $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$, the equation (A.12) can be further rewritten as

$$r_m(\phi_p) = \sqrt{A} x_0(\phi_p) e^{-j\Phi_0} \text{sinc}\left(D(\tilde{\theta}_m - \tilde{\phi}_p\right) \tag{A.13}$$

As a result, the array response for incident plane wave with physical angle ϕ_p on the m-th antenna element will be

$$\begin{aligned} a(\phi_p)[m] &= \frac{r_m(\phi_p)}{x_0(\phi_p)} = \sqrt{A} e^{-j\Phi_0} \text{sinc}\left(D(\tilde{\theta}_m - \tilde{\phi}_p\right) \\ &= \sqrt{A} e^{-j\Phi_0} \text{sinc}(D \cdot \frac{m}{D} - D\sin\phi_p) \\ &= \sqrt{A} e^{-j\Phi_0} \text{sinc}(m - D\sin\phi_p), \ \ \forall m \in [-\frac{M-1}{2}, \frac{M-1}{2}]. \end{aligned} \tag{A.14}$$

In the end, to keep the consistency with linear arrays, the DLA-based array response can also be represented via ordinal index $m \in [0, M-1]$ and Direction of Arrival (DoA) ϕ as

$$\begin{aligned} a(\phi)[m] &= \sqrt{A} e^{-j\Phi_0} \text{sinc}\left(m - \lfloor D \rfloor - D \cdot \sin\left(\frac{\pi}{2} - \phi\right)\right) \\ &\approx \sqrt{A} e^{-j\Phi_0} \text{sinc}\left(m - D - D \cdot \cos\phi\right) \\ &= \sqrt{A} e^{-j\Phi_0} \text{sinc}\left(m - D(1 + \cos\phi)\right), \ \ \forall m \in [0, M-1]. \end{aligned} \tag{A.15}$$

To this end, it is clear that DLA array serves as DFT operation to convert the received signal onto beamspace, which follows a 'sinc' function. The same characteristics is also shown in the spatial spectrum of received signal via ULA array [24], which emphasizes the similar properties between beamspace representation and spatial spectrum.

Appendix B

CRLB approximation in DLA-based channel tracking system

Here, we can easily see that Σ is the strength of channel response without the effect of β, while the channel response of DLA approximates the DFT of the channel response of ULA. It is thus inspired to apply DFT properties to deal with these complex summations to explore the close-form of CRLB. To be specific, we assume a steering vector of ULA

$$\mathbf{e}[\omega] = \left[1, e^{j2\pi\omega}, ..., e^{j2\pi(M-1)\omega}\right]^T, \tag{B.1}$$

where $\omega = \frac{d\cos\theta}{\lambda}$ and M is the total number of antennas. Then the M-th point DFT of $\mathbf{e}[\omega]$ will be

$$\begin{aligned}
\mathcal{E}_\omega[k] &= \sum_{m=0}^{M-1} e^{j2\pi m\omega} e^{-jmk\frac{2\pi}{M}} \\
&= \frac{1-e^{j2\pi M\left(\omega-\frac{k}{M}\right)}}{1-e^{j2\pi\left(\omega-\frac{k}{M}\right)}} \\
&= \frac{e^{j\pi M\left(\omega-\frac{k}{M}\right)}\left(-2j\sin\left(M\pi(\omega-\frac{k}{M})\right)\right)}{e^{j\pi\left(\omega-\frac{k}{M}\right)}\left(-2j\sin\left(\pi(\omega-\frac{k}{M})\right)\right)} \\
&\approx e^{j\pi(M-1)\left(\omega-\frac{k}{M}\right)}\mathrm{sinc}\left(\omega-\frac{k}{M}\right),
\end{aligned} \tag{B.2}$$

where DFT is defined as $\mathcal{X}_\omega[k] \triangleq \sum_{m=0}^{M-1} x[m]e^{-jmk\frac{2\pi}{M}}$. Recall the fact that $\|\mathbf{e}\|^2 = \|\mathcal{E}_\omega\|^2$ based on Parseval's theorem, the summation of a sequence of which each element is a square form of a sinc function can be derived as

$$\sum_{k=0}^{M-1} \mathrm{sinc}^2\left(\omega-\frac{k}{M}\right) = M. \tag{B.3}$$

Since the derivative of the sinc function appears in the other two summations in Equation (4.15), the partial derivative of $\mathcal{E}_\omega[k]$ attracts our attention

$$\begin{aligned}
\frac{\partial\mathcal{E}_\omega[k]}{\partial\omega} &= \sum_{m=0}^{M-1} j2\pi m e^{j2\pi m\omega} e^{-jmk\frac{2\pi}{M}} \\
&= j\pi(M-1)e^{j\pi(M-1)\left(\omega-\frac{k}{M}\right)}\mathrm{sinc}\left(\omega-\frac{k}{M}\right) + e^{j\pi(M-1)\left(\omega-\frac{k}{M}\right)}\frac{\partial\left(\mathrm{sinc}\left(\omega-\frac{k}{M}\right)\right)}{\partial\omega}.
\end{aligned} \tag{B.4}$$

Intuitively, Equation (B.4) is the DFT of another vector $\mathbf{f}$ which can be defined as

$$\mathbf{f} = \left[0, j2\pi e^{j2\pi\omega}, ..., j2\pi(M-1)e^{j2\pi(M-1)\omega}\right]^T. \tag{B.5}$$

Now, we have $\mathbf{e}$, $\mathcal{E}_\omega = DFT\{\mathbf{e}\}$, $\mathbf{f}$ and $\mathcal{F}_\omega = DFT\{\mathbf{f}\}$ where $\mathcal{F}_\omega = \frac{\partial \mathcal{E}_\omega[k]}{\partial \omega}$. According to the Plancherel theorem, denoting $\mathbf{e}^H\mathbf{f} = \mathcal{E}_\omega^H \mathcal{F}_\omega$, we have

$$\mathbf{e}^H\mathbf{f} = \sum_{m=0}^{M-1} j2\pi m = j\pi M(M-1), \tag{B.6}$$

$$\mathcal{E}_\omega^H \mathcal{F}_\omega = j\pi(M-1)\sum_{k=0}^{M-1} \operatorname{sinc}^2\left(\omega - \frac{k}{M}\right) + \sum_{k=0}^{M-1} \operatorname{sinc}\left(\omega - \frac{k}{M}\right)\frac{\partial\left(\operatorname{sinc}\left(\omega - \frac{k}{M}\right)\right)}{\partial\omega}. \tag{B.7}$$

By substitution with Equation (B.3), we can get

$$\sum_{k=0}^{M-1} \operatorname{sinc}\left(\omega - \frac{k}{M}\right)\frac{\partial\left(\operatorname{sinc}\left(\omega - \frac{k}{M}\right)\right)}{\partial\omega} = 0 \tag{B.8}$$

Last, we recall the Parseval's theorem again on $\mathbf{f}$, then we have $\|\mathbf{f}\|^2 = \|\mathcal{F}_\omega\|^2$ as follows

$$\|\mathbf{f}\|^2 = 4\pi^2 \sum_{m=0}^{M-1} m^2 = \frac{2\pi^2(M-1)M(2M-1)}{3}, \tag{B.9}$$

$$\|\mathcal{F}_\omega\|^2 = \pi^2(M-1)^2 \sum_{k=0}^{M-1} \operatorname{sinc}^2\left(\omega - \frac{k}{M}\right) + \sum_{k=0}^{M-1}\left(\frac{\partial\left(\operatorname{sinc}\left(\omega - \frac{k}{M}\right)\right)}{\partial\omega}\right)^2 + 0. \tag{B.10}$$

To this end, it is straight forward to derive the summation of a sequence of which each element is a square of the partial derivative of the sinc function

$$\begin{aligned}\sum_{k=0}^{M-1}\left(\frac{\partial\left(\operatorname{sinc}\left(\omega - \frac{k}{M}\right)\right)}{\partial\omega}\right)^2 &= \frac{2\pi^2(M-1)M(2M-1)}{3} - \pi^2 M(M-1)^2 \\ &= \frac{\pi^2\left(M^3 - M\right)}{3}.\end{aligned} \tag{B.11}$$

To sum up, based on two facts that the number of antenna elements is relatively large and the truncated channel vector preserves more than 80% of total energy, we can approximate

Φ, Ω and Σ based on above DFT properties as follows

$$\Phi \approx \frac{\pi^2 \left(M^3\right)}{3}, \tag{B.12}$$

$$\Omega \approx 0, \tag{B.13}$$

$$\Sigma \approx M. \tag{B.14}$$

Therefore, we can rewrite Equation (4.15) as

$$\mathbf{F} = \frac{2}{\sigma_e^2} \begin{bmatrix} \|\beta\|^2 \frac{\pi^2 M^3}{3} & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & M \end{bmatrix}, \tag{B.15}$$