Hybrid Beamforming for Millimeter Wave Communications

by

Jinlong Zhan
B.Sc., Xi'an University of Science and Technology, 2000
M.Sc., Xi'an University of Science and Technology, 2003

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Hybrid Beamforming for Millimeter Wave Communications

by

Jinlong Zhan
B.Sc., Xi'an University of Science and Technology, 2000
M.Sc., Xi'an University of Science and Technology, 2003

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Hong-chuan Yang, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Hong-chuan Yang, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

Communications over millimeter wave (mmWave) frequencies is a key component of the fifth generation (5G) cellular networks due to the large bandwidth available at mmWave bands. Thanks to the short wavelength of mmWave bands, large antenna arrays (32 to 256 elements are common) can be mounted at the transceivers. The array sizes are typical of a massive MIMO communication system, which makes fully digital beamforming difficult to implement due to high power consumption and hardware cost. This motivates the development of hybrid beamforming due to its versatile tradeoff between implementation cost (including hardware cost and power consumption) and system performance. However, due to the non-convex constraints on hardware (phase shifters), finding the global optima for hybrid beamforming design is often intractable. In this thesis, we focus on hybrid beamforming design for mmWave cellular communications both narrowband and wideband scenarios are considered.

Starting from narrowband SU-MIMO mmWave communications, we propose a Gram-Schmidt orthogonalization (GSO) aided hybrid precoding algorithm to reduce computation complexity. GSO is a recursive process that depends on the order in which the matrix columns are selected. A heuristic solution to the order of column selection is suggested according to the array response vector along which the full

digital precoder has the maximum projection. The proposed algorithm, not only constrained to uniform linear arrays (ULAs), can avoid the matrix inversion in designing the digital precoder compared to the orthogonal matching pursuit (OMP) algorithm.

For the narrowband MU-MIMO mmWave communications, we propose an interference cancellation (IC) framework on hybrid beamforming design for downlink mmWave multi-user massive MIMO system. Based on the proposed framework, three successive interference cancellation (SIC) aided hybrid beamforming algorithms are proposed to deal with inter-user and intra-user interference. Furthermore, the optimal detection order of data streams is derived according to the post-detection signal-to-interference-plus-noise ratio (SINR).

When considering wideband MU-MIMO mmWave communications, how to design a common RF beamformer across all subcarriers becomes the main challenge. Furthermore, the common RF beamformer in wideband channels leads to the need of more effective baseband schemes. By adopting a relaxation of the original mutual information and spectral efficiency maximization problems at the transceiver, we design the radio frequency (RF) precoder and combiner by leveraging the average of the covariance matrices of frequency domain channels, then a SIC aided baseband precoder and combiner are proposed to eliminate inter-user and intrauser interference.

Contents

List of Tables

List of Figures

ABBREVIATIONS

PCS	Personal Communication Services
FDMA	Frequency-Division Multiplexing Access
TDMA	Time-Division Multiplexing Access
CDMA	Code-Division Multiplexing Access
SMS	Short Message Service
MMS	Multimedia Messaging Service
LTE	Long Term Evolution
OFDM	Orthogonal Frequency-Division Multiplexing
MIMO	Multiple-Input Multiple-Output
IoT	Internet of Things
ITU	International Telecommunication Union
NR	New Radio
eMBB	Enhanced Mobile Broadband
mMTC	Massive Machine Type Communication
URLLC	Ultra-Reliable Low Latency Communication
mmWave	millimeter Wave
3GPP	Third Generation Partnership Project
ADC	Analog-to-Digital Converter
DAC	Digital-to-Analog Converter
PA	Power Amplifier
LNA	Low Noise Amplifier
RF	Radio Frequency
SIC	Successive Interference Cancellation
SU-MIMO	Single-User MIMO
GSO	Gram-Schmidt orthogonalization
OMP	Orthogonal Matching Pursuit
MU-MIMO	Multi-User MIMO
IC	Interference Cancellation
ZF	Zero-Forcing
BD	Block Diagonalization
EGT	Equal Gain Transmission
MMSE	Minimum Mean Square Error
SO	Successive Optimization
SVD	Singular Value Decomposition

MATHEMATICAL NOTATIONS

$\mathbf{A}^{\dagger}$	Moore–Penrose inverse of Matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of Matrix $\mathbf{A}$
$\mathbf{A}^{H}$	Conjugate transpose of Matrix $\mathbf{A}$
$\mathbf{A}^{T}$	Transpose of Matrix $\mathbf{A}$
$\|\mathbf{A}\|_F$	Frobenius norm of Matrix $\mathbf{A}$
$\|\mathbf{A}\|$	Determinant of Matrix $\mathbf{A}$
argmax	Argument of the maximum
argmin	Argument of the minimum
max	Maximum value of a function
min	Minimum value of a function
$\mathbf{A}(i,:)$	The i-th row of Matrix $\mathbf{A}$
$\mathbf{A}(:,j)$	The j-th column of Matrix $\mathbf{A}$
$(\mathbf{A})_{i,j}$	The element in the i-th row, the j-th column of matrix $\mathbf{A}$
$\mathrm{Tr}(\mathbf{A})$	Trace of Matrix $\mathbf{A}$
$\mathbb{E}\{\cdot\}$	Mathematical expectation
$\angle\mathbf{A}$	Phase of all elements in Matrix $\mathbf{A}$
$\mathbf{I}_n$	Identity matrix with dimension $n \times n$
$\mathcal{CN}(0,\mathbf{I})$	Zero-mean complex Gaussian distribution with covariance matrix $\mathbf{I}$
$\mathbb{C}^{m\times n}$	$m \times n$ dimensional complex space

ACKNOWLEDGEMENTS

First and foremost I am extremely grateful to my supervisors, Prof. Xiaodai Dong for her invaluable advice, continuous support, and patience during my PhD study. Her immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

Furthermore, I also would like to express my sincere thanks to Prof. Hong-chuan Yang, Prof. Yang Shi and Prof. Julian Cheng as the members in my supervisory committee. Their valuable comments and suggestions have greatly improved my dissertation. I am also appreciative of Prof. Wu-Sheng Lu and Prof. Tao Lu, for many valuable talks and generous help they have provided for my research and thesis. I am also deeply grateful to all of the members of staff in the department of Electrical and Computer Engineering, particularly Dan Mai, Ashleigh Carlsen, Jessica Fox for their continuous encouragement and support.

Moreover, I would like to thank all the team members, particularly Yiming Huo, Yizhou Zhu, Minh Tu Hoang, Yu Zhang, Xiang Chen, Amy Sun, Baojun Hu, Xin Li, Bo Zhao, Yu Zhang, Haizhou Bao, Erfan Khorram, Le Hung Nguyen, Amir Mehdirezaei. It is their kind help and support that have made my study and life in Victoria a wonderful time.

I would also like to take this opportunity to thank my colleagues at Xi'an University of Posts and Telecommunications for their constant help and encouragement.

Finally, I would like to express my gratitude to my parents, my father-in-law, my wife and my son whose constant love and support keep me motivated and confident. Without their tremendous understanding and encouragement in the past few years, it would be impossible for me to complete my study.

Jinlong Zhan, Victoria, BC, Canada

DEDICATION

To my family

Chapter 1

Introduction

More than a century ago, in 1895, the capability of using electromagnetic waves to transmit signals wirelessly was successfully demonstrated, for the first time, in the famous wireless telegraphy experiment conducted by Guglielmo Marconi [3], and since then new wireless communications methods and services have been enthusiastically adopted by people throughout the world. In this dissertation, we will concentrate on cellular mobile communications, not only because they are of great current interest but also because the features of many other wireless systems can be easily understood as special cases or simple generalizations of the features of cellular mobile communications [4]. Cellular mobile communications have experienced exponential growth over the last few decades and there are currently about seven billion users worldwide, including about six billion smartphone users [5]. A brief history of the evolution of cellular mobile communications throughout the world is useful in order to appreciate the enormous impact that cellular radio and personal communication services (PCS) will have on all of us over the next several decades [4, 6–11]. Figure 1.1 shows the evolution of cellular mobile communications, and we will discuss the brief history of the evolution of cellular mobile communications in the following section.

1.1 Evolution of Cellular Mobile Communications

1.1.1 1G: Where it all began

The first generation of mobile networks (1G) as they were retroactively dubbed when the next generation was introduced – was launched by Nippon Telegraph and Telephone (NTT) in Tokyo in 1979. By 1984, NTT had rolled out 1G to cover the

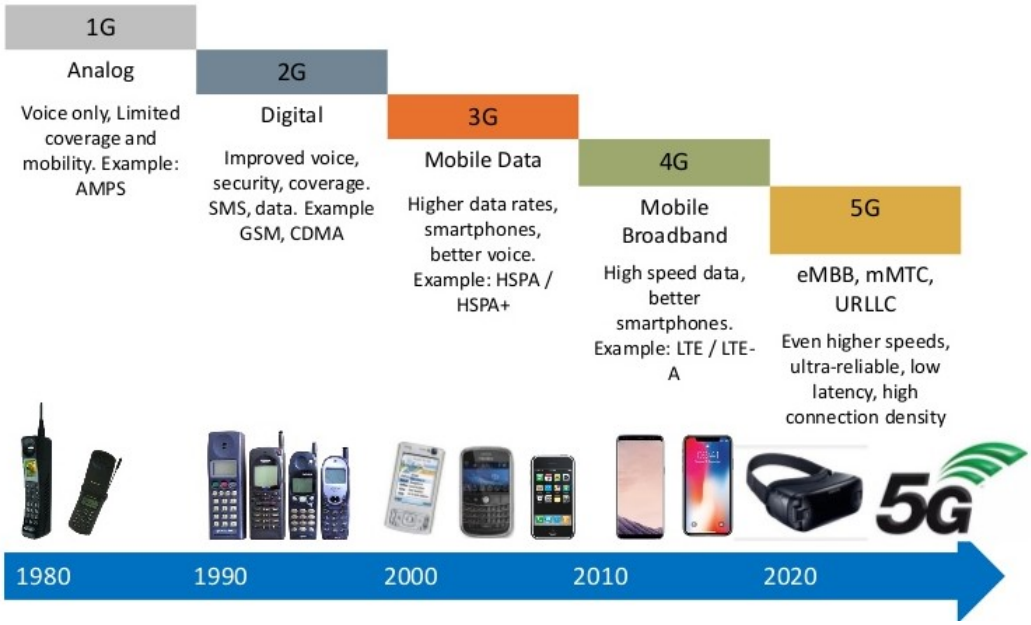

Figure 1.1: Evolution of Wireless Communication [1].

whole Japan. In 1983, the US approved the first 1G operations and the Motorola's DynaTAC became one of the first 'mobile' phones to see widespread use stateside. Other countries such as Canada and the UK rolled out their own 1G networks a few years later. China deployed and commercialized its 1G mobile communication system in Guangdong on November 18, 1987. The 1G mobile communications were based on analog communications by using the frequency-division multiplexing access (FDMA). It only allowed voice signal transfer with limited and unstable spatial coverage [5]. There was no roaming support between various operators and, as different systems operated on different frequency ranges, there was no compatibility between systems. Worse of all, calls weren't encrypted, so anyone with a radio scanner could drop in on a call.

1.1.2 2G: The Digital Revolution

The second generation of mobile networks (2G), was launched under the Global System for Mobile Communications (GSM) standard in Finland in 1991. The 2G mobile communications use digital communications where the time-division multiplexing access (TDMA) and code-division multiplexing access (CDMA) were adopted. For the first time, calls could be encrypted and digital voice calls were significantly clearer with less static and background crackling. Furthermore, people could send text mes-

sages (Short Message Service, SMS), picture messages, and multimedia messages (Multimedia Messaging Service, MMS) on their phones. The transfer speeds of 2G mobile communications were initially only around 9.6 Kbps. By the end of the era, speeds of 40 Kbps were achievable and Enhanced Data rates for GSM Evolution (EDGE) connections offered speeds of up to 500 Kbps.

1.1.3 3G: The 'Packet-Switching' Revolution

The third generation of mobile networks (3G) was launched by NTT DoCoMo in Japan in 2001 and aimed to standardize the network protocol used by vendors. This meant that users could access data from any location in the world as the 'data packets' that drive network connectivity were standardized. This made international roaming services a real possibility for the first time. The 3G mobile communications employ variations of advanced CDMA techniques and have 4 times the data transferring capabilities reaching up to 2 Mbps on average compared to 2G. The increased data transfer capabilities led to the rise of new services such as video conferencing, video streaming, and voice over IP (such as Skype).

1.1.4 4G: The Streaming Era

The forth generation of mobile networks (4G) was first deployed in Sweden and Norway in 2009 as the Long Term Evolution (LTE) 4G standard. With the help of orthogonal frequency-division multiplexing (OFDM) and multiple-input multiple-output (MIMO) techniques, 4G, including the LTE and LTE-advanced (LTE-A) was developed. They are able to offer a dramatically faster speed than 3G, providing a data rate of tens of megabytes per second. It was subsequently introduced throughout the world and made high-quality video streaming a reality for millions of consumers. The milestone of the 4G era was the prosperous widespread usage of smartphones worldwide, which changed the lifestyle of human beings and the way people communicate with each other. In terms of the frequency spectrums that are designated for the different generations of mobile communications, we can make two observations. First, more frequency bands have been gradually released for a larger channel bandwidth that can meet the demands for higher data rates. Second, all the released frequency bands are below 6 GHz (Sub-6 GHz), which is overcrowded as shown in Figure 1.2.

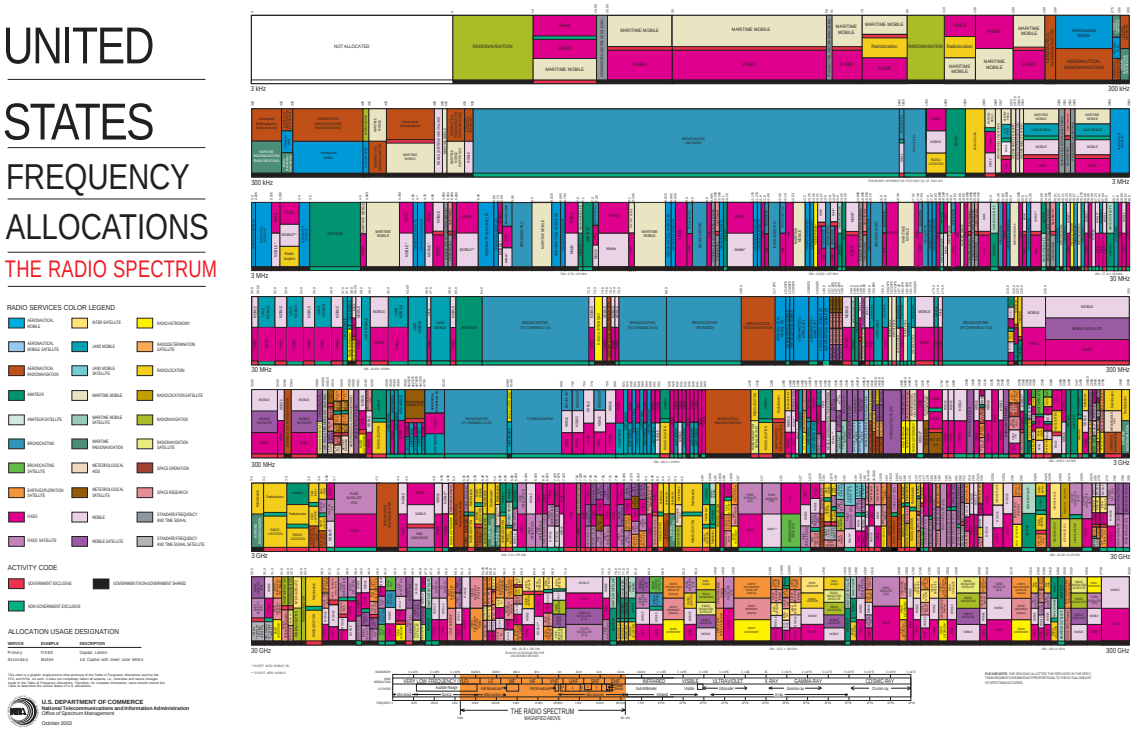

Figure 1.2: United States Frequency Allocation [2].

1.1.5 5G: The Internet of Things Era

With the rapid growing of the number of consumer wireless devices in use and the expansion of the Internet of Things (IoT), the amount of mobile data transfer is almost doubled every year, which results the 4G mobile network infrastructure can no longer meet the needs for high-speed wireless data transmission. Therefore, from the second decade of the 21st century, the fifth-generation (5G) of mobile communications emerges with the outlook to the sixth-generation (6G) [12–14]. The 5G has been first commercially deployed in 2019 in South Korea and is on the corner of massive commercialization. The 5G is expected to support a data rate of a few Gbps, which is much more than 4G can support. The International Telecommunication Union (ITU) has defined three major application scenarios for 5G New Radio (NR): the enhanced mobile broadband (eMBB), massive machine type communication (mMTC), and ultra-reliable low latency communication (URLLC). Importantly, from the frequency resource perspective, the uniqueness of 5G, in comparison with 3G and 4G, is the utilization of millimeter-wave (mmWave) frequencies from 30 GHz to 300 GHz. The mmWave band can provide a much wider bandwidth than the

current centimeter waveband, which is shown in Figure 1.2. The Third Generation Partnership Project (3GPP) has divided the 5G NR into Frequency Ranges 1 (FR1) band, i.e., 410 – 7125 MHz, and FR2 band, or also called mmWave band, i.e., 24.25 – 52.6 GHz [15]. Furthermore, a shorter wavelength at mmWave bands enables a large number of antennas to be mounted at the transceivers, which facilitates the implementation of massive MIMO [16], [17]. It has been theoretically proved that massive MIMO can achieve orders of magnitude improvement in spectral efficiency with simple linear precoding [18–20]. The above two advantages provide mmWave communications the tremendous potential ability to support multiple gigabit-per-second transmission. Besides cellular communications [21–23], 5G has been considered for many other applications including device-to-device communications [24, 25], and vehicular communications [26–28].

1.2 Beamforming

1.2.1 Analog Beamforming

Analog beamforming is an approach that relies entirely on RF domain processing to steer the transmit or receive beams in the desired directions [18], [29], [30], [31]. The configuration of analog beamforming is illustrated in Figure 1.3. In this configuration, several antenna elements are connected via phase shifters to a single RF chain, the phase shifter weights are adaptively adjusted using a specific strategy to steer the beam and meet a given objective, for example to maximize received signal power. However, analog beamforming with a single beamformer only supports single-user and single-stream transmission, which means it is not possible to realize multi-stream or multi-user benefits associated with MIMO.

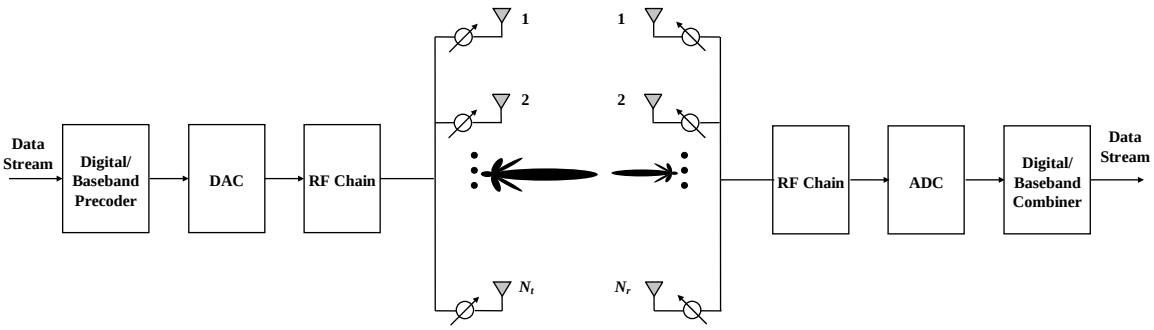

Figure 1.3: Analog beamforming architecture for MIMO sytems.

1.2.2 Digital Beamforming

In conventional cellular systems working at sub-6GHz cellular frequencies like the Institute of Electrical and Electronics Engineers (IEEE) 802.16m, 3GPP LTE, and LTE-A, the number of transmit and receive antennas is small. For example, IEEE 802.16m and 3GPP LTE-A target MIMO schemes for the same sets of antenna configurations: 2, 4, or 8 transmit antennas and a minimum of 2 receive antennas in the downlink (DL), and 1, 2, or 4 transmit antennas in the uplink (UL) with a minimum of 2 receive antennas [32]. In this case, precoding and combining are performed in the baseband using fully digital signal processing, which is also called fully digital beamforming. This digital processing allows better control over the entries of the precoding and combining matrices, which in turn facilitates the implementation of sophisticated single-user, multi-user, and multi-cell beamforming algorithms [33]. The configuration of fully digital beamforming is illustrated in Figure 1.4. In this configuration, each antenna is connected to a dedicated RF chain which includes the analog-to-digital converters (ADCs), digital-to-analog converters (DACs), signal mixer, and power amplifier (PA) or the low noise amplifier (LNA). Recently, a Q-band 64-channel fully digital beamforming transceiver for 5G communications has been proposed and implemented, covering a frequency range from 37 to 42.5 GHz [34]. However, the cost of hardware implementation and the burden of signal processing in the baseband will increase quickly as the antenna number increases or the channel bandwidth broadens, limiting its commercialization [11], which will be discussed in the following subsection.

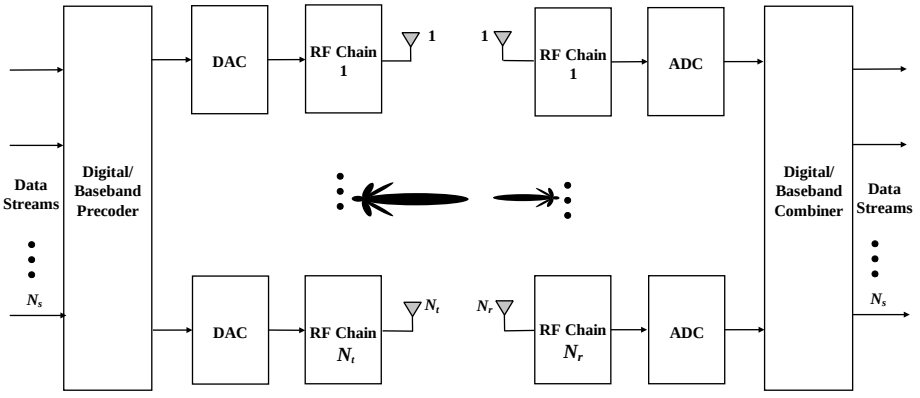

Figure 1.4: Conventional digital beamforming architecture for MIMO sytems.

1.2.3 Hybrid Beamforming

For mmWave communication systems, large antenna arrays (32 to 256 elements are common) can be mounted at the transceivers. The array sizes are typical of a massive MIMO communication system, which makes fully digital beamforming difficult to implement due to several hardware constraints [35,36]. First, the practical implementation of the PA or LNA, the RF chain associated with each antenna element and all baseband connection is difficult at mmWave bands [18], [37], [38], because these devices have to be packed behand each antenna, and all the antenna elements are placed close to each other to avoid granting lobes. Although the distances between antenna elements at mmWave communications can be smaller than that at lower frequency systems, the space limitation is still an issue along with the increasing of antenna elements. This space limitation prevents using a complete RF chain per antenna. Second, power consumption is also a limiting factor: (i) PAs and ADCs are power hungry devices especially at mmWave bands [22], [39], [40]. For example, according to the data from a number of recent papers proposing prototype devices, The power consumption of a PA is as much as 40-250 mW [41–43], the power consumption of an LNA is as much as 4-86 mW [44–48], and the power consumption of an ADC is as much as 15-795 mW [49–52]. (ii) a digital conversion stage per antenna leads to a large demand on digital signal processing, since many parallel Giga samples per second data streams have to be processed, with excessive power consumption as well [83]. Therefore, having a separate RF chain and data converter for every antenna element can drive the cost and power beyond allocated design budgets, which results in fully digital beamforming difficult to implement at mmWave bands. To address this issue, a hybrid beamforming structure (including hybrid precoding and combining) which can be implemented with only a small number of RF chains has been extensively investigated in recent years due to its potential low cost and energy efficiency [53–56]. It has been shown that such an architecture can approach the performance of a fully digital scheme if the number of RF chains is twice the total number of data streams [56] [57]. Hybrid precoding design can be divided into the following categories, matrix decomposition [56,58,59], two-stage based design [60–62], alternating minimization [63], successive interference cancellation (SIC) [64], [65] and deep-learning based [66–70]. Hybrid beamforming will be discussed in a greater detail in the upcoming chapters. The configuration of hybrid beamforming is illustrated in Figure 1.5. In this configuration, hybrid beamforming decouples the traditional

full digital precoder into a low-dimension baseband (or digital) precoder and a high-dimension RF (or analog) precoder, and the RF precoder is implemented via phase shifters.

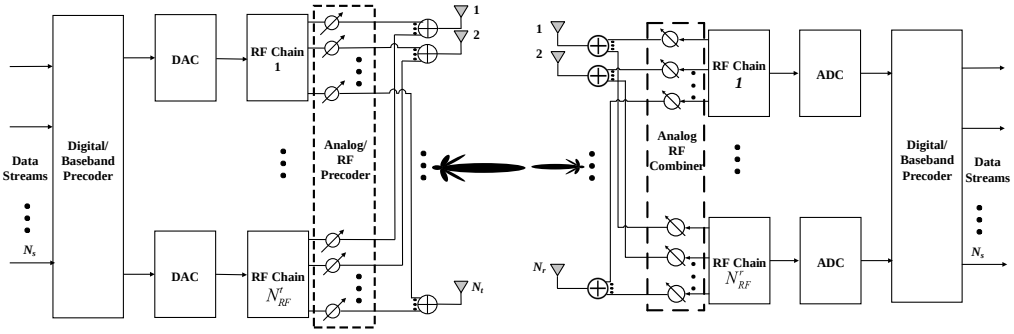

Figure 1.5: Hybrid beamforming architecture for MIMO sytems.

1.3 Contribution and Organization of the Thesis

The organization of the thesis is shown in Figure 1.6. First, in Chapter 2, the propagation characteristics of mmWave bands and mmWave channel models are introduced.

Second, in Chapter 3, for narrowband mmWave channel with single-user MIMO (SU-MIMO) transmission, we focus on how to further reduce the computation complexity when designing digital beamforming and achieve spectral efficiency close to that obtained by full digital solutions. We propose a novel hybrid precoding algorithm based on Gram-Schmidt orthogonalization (GSO) which can avoid the matrix inversion in designing the digital precoder compared to Orthogonal Matching Pursuit (OMP). Since GSO is a recursive process that depends on the order of the vectors to be orthogonalized. We propose a heuristic solution to the order of columns selection to perform GSO.

Chapter 4 contributes to designing hybrid beamforming narrowband mmWave channel with multi-user MIMO (MU-MIMO) transmission. In this scenario, the inter-user interference and intra-user (inter-stream) interference cancellation are our main concerns. Instead of using linear algorithms to cancel inter-user and intra-user interference, we propose a new framework, that is, an interference cancellation (IC)-aided hybrid beamforming design framework for downlink mmWave multi-user massive MIMO systems, in which non-linear algorithms can be used to suppress inter-user

and intra-user interference. Under the proposed framework, three successive interference cancellation (SIC)-aided hybrid beamforming algorithms that maximize the sum spectral efficiency are presented according to the dominance of inter-user interference, intra-user interference, or both. To deal with error propagation in the SIC procedure, we propose the optimal detection order of data streams according to the post-detection signal-to-interference-plus-noise ratio (SINR), which can further improve the spectral efficiency performance of the proposed three algorithms. The proposed algorithms are illustrated to outperform the state of the art hybrid beamforming designs and can achieve good spectral efficiency performance even with finite resolution phase shifters and channel estimation error. In addition, the proposed algorithms have closed-form solutions with relatively low complexity.

Chapter 5 further considers hybrid beamforming design for wideband mmWave multi-user MIMO-OFDM Systems. How to design a common RF beamformer across all subcarriers and consequent more effective baseband schemes are our main concerns. We develop a near-optimal closed-form RF beamforming solution for wideband mmWave MU-MIMO OFDM systems. In our design, we adopt a relaxation of the original mutual information maximization problem at the transmitter and a relaxation of the spectral efficiency maximization problem at the receiver. For the relaxed problems, we design RF precoder and combiner which have almost the same spectral efficiency as the original problems. We propose an SIC-aided baseband beamforming algorithm on a per-subcarrier basis to cancel both inter-user and intra-user interference. Specifically, we use successive optimization (SO)-based SIC to cancel inter-user interference and Zero Forcing (ZF) SIC to cancel intra-user interference. The optimal detection order of data streams is derived according to the post-detection SINR. Furthermore, We analyze the performance of the proposed algorithm in the presence of channel estimation error. By modelling the estimation error as independent complex Gaussian random variables and uncorrelated with the ideal channel, tight approximations for both the original mutual information and spectral efficiency are derived. Numerical results show that the proposed algorithm outperforms the state of the art hybrid beamforming designs and can achieve good spectral efficiency performance even with finite resolution phase shifters and channel estimation error.

In the end, Chapter 6 summarizes the entire thesis and proposes the future work.

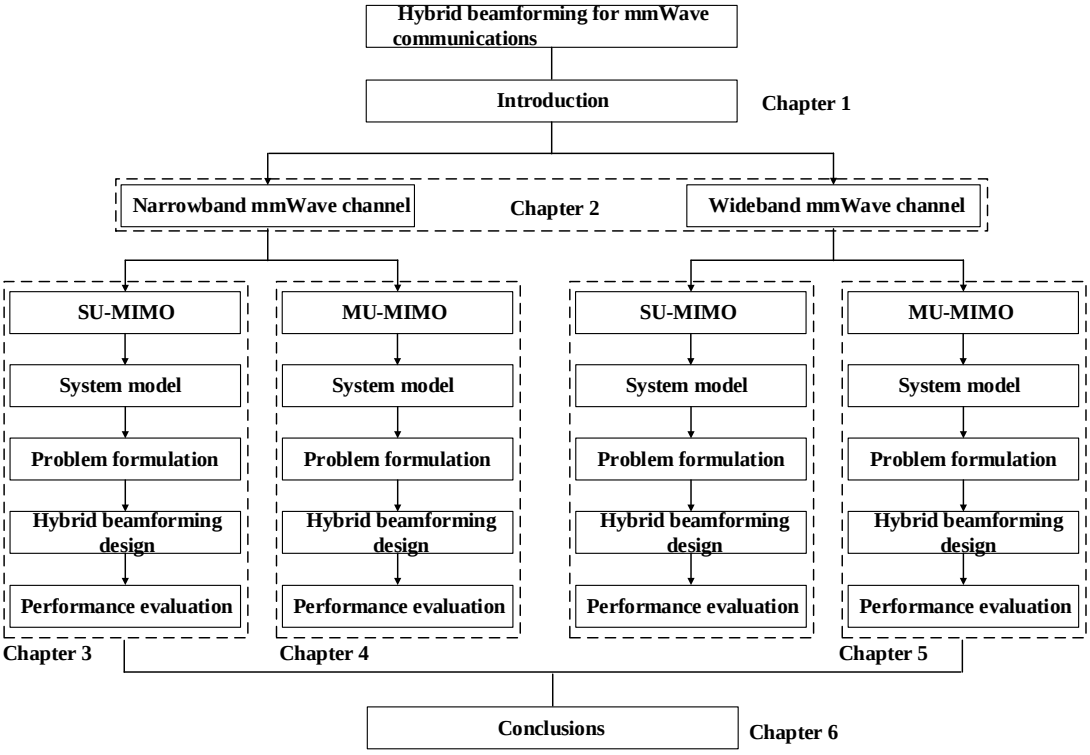

Figure 1.6: Organization of the thesis.

Chapter 2

mmWave Channel Model

In this chapter, we will introduce propagation characteristics of mmWave band and mmWave channel models.

2.1 Propagation Characteristics of mmWave

Due to the increase in frequency, mmWaves have different propagation characteristics compared to microwaves with frequencies below 6 GHz. For example, mmWaves cannot efficiently penetrate and diffract around obstacles, e.g., vehicles, buildings and human bodies. This makes mmWaves propagation is characterized by few diffracting MPCs and high path loss. mmWaves propagation is illustrated in Figure 2.1. The rays transmit from the transmitter (Tx), propagate through the channel which is represented by N_c clusters, and finally arrive at the receiver (Rx). In the spherical coordinates, the angle-of-departure (AoD) and angle-of-arrival (AoA) at both TX and RX sides are used to characterize the rays. Furthermore, a cluster generally consists of a few rays, which come from the same scatterer (e.g., a group of buildings), with similar characteristics, e.g., delays, AoDs and AoAs. In the following subsections, we focus on several main channel properties in mmWave bands and introduce mmWave channel models.

2.1.1 Attenuation and Blockage

During propagation, mmWaves may be partially or totally absorbed by the propogation medium, which results in attenuation. Thus the attenuation due to molecular absorption should be considered in mmWave systems, although this is not a con-

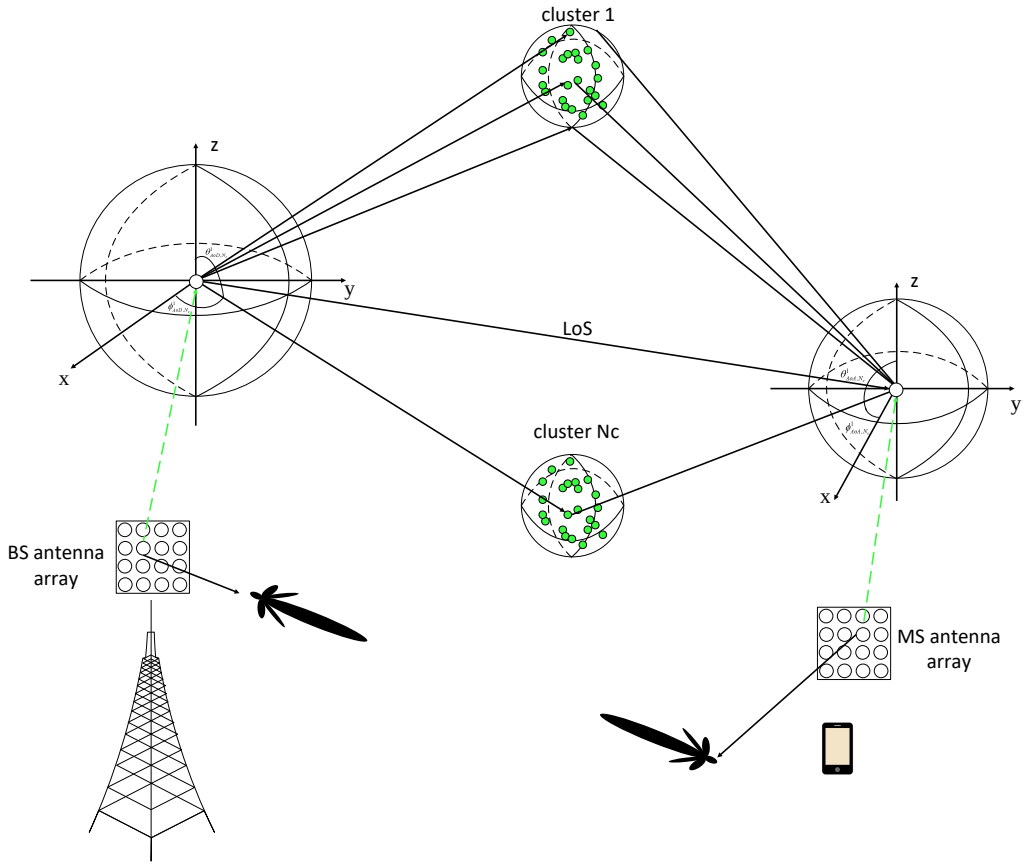

Figure 2.1: Illustration of mmWave propagation.

cerned problem in microwave systems [71]. For example, the attenuation due to gas absorption for a 60 GHz waveform is more than 10 dB/km, while a 700 MHz waveform experiences an attenuation on the order of 0.01 dB/km. Additionally, a more significant challenge is mmWaves's vulnerability to blockage [18]. Materials such as brick can attenuate mmWave signals by as much as 40 to 80 dB [21], [72], [73], outdoor tinted glass can attenuate mmWave signals by 40.1 dB at 28 GHz [73], and the human body itself can result in a 20 to 35 dB loss [74]. Furthermore, Foliage loss can also be significant [75], [76].

2.1.2 Channel Sparsity

mmWave channels are expected to have limited scattering [77], [78], [79], which results in sparsity in the angle and delay domains. For example, only up to five spatial lobes are found [80] in dense-urban non-line-of-sight (NLOS) environments, and the

delay/angle spreading within each cluster is relatively small.

2.1.3 Large Bandwidth and Large Antenna Arrays

As mentioned in Chapter 1, the main motivation for shifting cellular communication to mmWave bands is to make use of the large bandwidth available at high frequencies. This enables users to be assigned a large communication channel bandwidth, and to send data at high rates. Furthermore, smaller wavelengths make large antenna arrays (32 to 256 elements are common) mounted at the transceivers feasible, which facilitates the implementation of massive MIMO, but still occupy the same or even a smaller physical size [20], [33]. Therefore, the channel model should consider high resolution in both delay and angular domains. Moreover, various types of antenna arrays, such as the uniform linear array (ULA) and uniform planar array (UPA), are both considered.

2.1.4 Random Cluster Numbers

For the mmWave bands, the cluster/time-cluster numbers are small and random, and are well-modeled by a Poisson distribution [71], [79], [80]. In [81], the mean cluster number is 12 while it is less than 4 in [79] (Note that the definition of cluster is different in these two references). Some channel properties, e.g., capacity, will change correspondingly along with the random cluster numbers.

2.2 mmWave Channel Model

2.2.1 Narrow Band mmWave Channel Model

According to the propagation characteristics of mmWave bands, mmWave channels are expected to have limited scattering capabilities with only a few scattering clusters [72], [82], [83]. Therefore, we adopt a narrow band flat fading clustered channel representation, which is based on the extended Saleh-Valenzuela model [58]. Using the clustered channel model, the matrix channel $\mathbf{H}$ is assumed to be a sum of the contributions of N_c scattering clusters, each of which contribute N_p propagation paths to the channel matrix $\mathbf{H}$. The narrow band flat fading channel matrix $\mathbf{H}$ can be

written as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_C N_P}} \sum_{i=1}^{N_C} \sum_{l=1}^{N_P} \alpha_{il} \mathbf{a}_{MS}(\theta_{il}) \mathbf{a}_{BS}(\varphi_{il})^H, \tag{2.1}$$

where $\alpha_{il}, i = 1, 2 \cdots N_c, l = 1, 2 \cdots N_P$ is the complex gain of the l-th ray in the i-th cluster, which follows $\mathcal{CN}(0, 1)$, while $\mathbf{a}_{\mathrm{MS}}(\theta_{il})$ and $\mathbf{a}_{\mathrm{BS}}(\varphi_{il})$ are the MS and BS array response vectors, respectively, where θ_{il} and φ_{il} denote the AoA and AoD of the l-th ray in the i-th cluster, respectively. The channel matrix $\mathbf{H}$ can be further represented

$$\begin{aligned} \mathbf{H} &= \underbrace{\left[\mathbf{a}_{\mathrm{MS}}(\theta_{11}), \cdots, \mathbf{a}_{\mathrm{MS}}(\theta_{N_c N_p})\right]}_{\mathbf{A}_{\mathrm{r}}} \underbrace{\mathrm{diag}\left[\begin{array}{cccc} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{N_c N_p} \end{array}\right]}_{\mathbf{D}} \\ &\quad \times \underbrace{\left[\mathbf{a}_{\mathrm{BS}}{}^H(\varphi_{11}), \cdots, \mathbf{a}_{\mathrm{BS}}{}^H(\varphi_{N_c N_p})\right]^T}_{\mathbf{A}_{\mathrm{t}}^H} \\ &= \mathbf{A}_{\mathrm{r}} \mathbf{D} \mathbf{A}_{\mathrm{t}}^H, \end{aligned} \tag{2.2}$$

where $\mathbf{D} = \mathrm{diag}\left[\begin{array}{cccc} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{N_c N_p} \end{array}\right]$ denotes complex gain matrix, whose (i, l)-th element α_{il} represents the complex gain of the l-th ray in the i-th cluster. $\mathbf{A}_{\mathrm{r}} = \left[\mathbf{a}_{\mathrm{MS}}(\theta_{11}), \cdots, \mathbf{a}_{\mathrm{MS}}(\theta_{N_c N_p})\right]$ and $\mathbf{A}_{\mathrm{t}} = \left[\mathbf{a}_{\mathrm{BS}}(\varphi_{11}), \cdots, \mathbf{a}_{\mathrm{BS}}(\varphi_{N_c N_p})\right]$ are the MS and BS array response matrix, respectively. For an N-element ULA, the array response vector is given by

$$\mathbf{a}_{\mathrm{ULA}}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi}{\lambda}\Delta d \sin(\theta)}, \cdots, e^{j\frac{2\pi}{\lambda}(N-1)\Delta d \sin(\theta)}\right]^T, \tag{2.3}$$

where λ is the wavelength of the carrier, and Δd denotes the antenna element spacing. In the UPA case, assume an (M, N)-element UPA in the xy-plane, the array response vector can be derived according to [84].

$$\begin{aligned} \mathbf{a}_{\mathrm{UPA}}(\phi, \theta) = \frac{1}{\sqrt{MN}} \Big[1, \cdots, e^{j\frac{2\pi\Delta d}{\lambda}(m \cos\phi \sin\theta + n \sin\phi \sin\theta)}, \\ \cdots, e^{j\frac{2\pi\Delta d}{\lambda}[(M-1)\cos\phi \sin\theta + (N-1) n \sin\phi \sin\theta]}\Big]^T, \end{aligned} \tag{2.4}$$

where m and n are the antenna element indexes with $0 \leq m \leq M-1, 0 \leq n \leq N-1$.

2.2.2 Wideband mmWave Channel Model

When considering wideband mmWave communications, we adopt a clustered wideband mmWave channel model with N_c clusters to incorporate the wideband and limited scattering characteristics of mmWave channel [78], [85], [86]. Each cluster c

Table 2.1: Channel parameters

Parameters	**Description**
N_t	Number of BS antennas
N_r	Number of MS antennas
K	Number of subcarriers
N_c	Number of cluster
N_p	Number of path for each cluster
τ_c	Time delay of cluster c
α_{cl}	Complex gain of cluster c path l
T_s	Sampling period
d	Time delay
D	Cyclic prefix length
θ_c	AoA of cluster c
γ_c	AoA shift of cluster c
φ_c	AoD of cluster c
ϕ_c	AoD shift of cluster c
Δd	Antenna element spacing
$\mathbf{a}_{MS}(\theta_c)$	MS array response vector
$\mathbf{a}_{BS}(\varphi_c)$	BS array response vector

has a time delay $\tau_c \in \mathbb{R}$ and is further assumed to contribute with N_p rays/paths between BS and MS. Moreover, let $g(\tau)$ represent the path-shaping function for T_s-spaced signaling evaluated at τ seconds. According to this model, the delay-d MIMO channel matrix between the MS and the BS, $\mathbf{H}[d]$ can be written as [85], [86]

$$\mathbf{H}[d] = \sqrt{\frac{N_t N_r}{N_C N_P}} \sum_{c=1}^{N_C} \sum_{l=1}^{N_P} \alpha_{cl} g(dT_s - \tau_c - \tau_l) \mathbf{a}_{MS}(\theta_c - \gamma_l) {\mathbf{a}_{BS}}^H(\varphi_c - \phi_l), \tag{2.5}$$

where $\alpha_{cl}, c = 1, 2 \cdots N_c, l = 1, 2 \cdots N_P$ is the complex gain of the l-th ray in the c-th cluster, which follows $\mathcal{CN}(0, 1)$, while $\mathbf{a}_{\text{MS}}(\theta_c)$ and $\mathbf{a}_{\text{BS}}(\varphi_c)$ are the MS and BS array response vectors, respectively, θ_c and φ_c denote the AoA and the AoD of the c-th cluster, respectively. γ_l and ϕ_l denote relative AoA and AoD shift of the l-th ray, respectively. Given the delay-d MIMO channel model, the frequency channel response at subcarrier k, $\mathbf{H}[k]$, can be then expressed as

$$\mathbf{H}[k] = \sum_{d=0}^{D-1} \mathbf{H}[d] e^{-j\frac{2\pi k}{K} d}, \tag{2.6}$$

where D is the length of cyclic prefix. For clarity, we summarize all the channel parameters in Table 2.1.

Chapter 3

Gram-Schmidt Orthogonalization Aided Hybrid Precoding in Millimeter-Wave Massive MIMO Systems

3.1 Introduction

In this chapter, we consider hybrid beamforming for point-to-point (single-user) mmWave massive MIMO communication systems. We focus on how to reduce the computation complexity of the hybrid beamforming.

3.1.1 Prior Works and Motivations

Hybrid beamforming for point-to-point mmWave massive MIMO communication systems has been investigated in the recent literature. In [56], the authors proposed a heuristic hybrid precoding algorithm that first calculates the optimal baseband precoder assuming a fixed analog precoder and then calculates a local optimal analog precoder by an iterative coordinate descent algorithm. However, the digital precoder is required to be a scaled unitary matrix. In [58], by leveraging the spatially sparse characteristic of a mmWave channel, the spectral efficiency maximization problem can be formulated as minimizing the Frobenius norm between the full digital beamforming and the overall hybrid beamforming. Then a sparse signal reconstruction algorithm via OMP was proposed to find the optimal solution. The OMP algorithm requires

the computation of matrix inversion iteratively when designing the digital precoder utilizing least square (LS) algorithm, thus leading to high computation complexity. In order to improve the computation efficiency of the OMP algorithm, several low complexity hybrid precoding solutions were proposed [87–90]. An Orthogonal Based Matching Pursuit (OBMP) algorithm was proposed in [87], which utilizes the DFT matrix (in lieu of array response matrix in OMP) as a set of candidate analog precoder. Due to the orthogonality of DFT matrix, the computation of matrix inversion can be avoided when designing the digital precoder. However, there is performance loss in the OBMP algorithm, and it is only applicable to ULA. In [88] and [89], the authors proposed several low complexity hybrid beamforming solutions which provide different trade-offs between performance and complexity. However, the digital precoder is required to be a scaled unitary matrix, and the number of data streams equals to the number of RF chains. Using a complex Hadamard matrix as a set of candidate analog precoder was proposed in [90], which is actually equivalent to quantizing the phase of RF precoder matrix [60]. Minimizing the previously mentioned Frobenius norm can also be solved via matrix decomposition [91] or matrix factorization [92]. In [93], based on the matrix-monotonic optimization theory, the combination of the optimal analog precoder and combiner was proved equivalent to eigenchannels selection for various optimal hybrid MIMO transceivers. In [63], an alternating minimization (AltMin) algorithm based on manifold optimization was proposed for the fully-connected structure. For the partially-connected structure, an AltMin algorithm based on semidefinite relaxation (SDR) is presented by offering optimal solutions for both subproblems of analog and digital precoders in each alternating iteration. In [94], using the minimum mean squared error (MMSE) criterion, the authors proposed a manifold optimization-based hybrid beamforming algorithm, which, however, suffers from high computation complexity. In [95], based on the alternating minimization of approximation gap framework, the authors proposed a minimal gap iterative quantization algorithm, which results in low complexity and low mean squared error (MSE). Moreover, in some existing hybrid beamforming designs mentioned previously, the hybrid receive combiner was optimized by the MMSE criterion [56], [58], [96]. In [64], the authors proposed a hybrid precoding for fixed partially connected structure based on SIC. Given that the sub-arrays can be dynamically constructed, the dynamic sub-array partitioning hybrid precoding was proposed in [86]. However, it requires an exhaustive search for all sub-array constructions to find the optimal solution.

Furthermore, a Gram-Schimidt orthogonalization (GSO) [97] based hybrid precoding method was proposed for frequency selective fading mmWave MIMO systems [85] and [98], which modifies the OMP algorithm by adding a Gram-Schmidt orthogonalization step in each iteration to project the candidate beamforming codewords on the orthogonal complement of the subspace spanned by the selected codewords in RF precoder matrix. However, the method proposed in [85] and [98] is intuited by the effective channel seen through the RF precoders lens will have better coverage over the dominant subspaces belonging to the actual channel matrix if the RF beamforming vectors to be orthogonal (or close to orthogonal). Furthermore, the method proposed in [85] and [98] requires singular value decomposition (SVD) twice and matrix inversion in designing the digital precoder, which increases computational complexity significantly. Moreover, the method proposed in [85] and [98] cannot be applied to partially connected hybrid precoding mmWave massive MIMO systems with each sub-array connected to one RF chain because the RF precoder of each sub-array is a vector in this scenario.

Motivated by the aforementioned considerations, we are aiming to design hybrid precoding and combining algorithm which can further reduce the computation complexity by avoiding matrix inversion when designing digital beamforming and achieve spectral efficiency close to that obtained by full digital solutions. In this chapter, we propose a low-complexity hybrid precoding algorithm based on GSO, assuming perfect and instantaneous channel state information (CSI) at the base station (BS) and the mobile station (MS). The rest of this chapter is organized as follows. In Section 3.2, we introduce mmWave hybrid precoding and combining system model, both fully connected and partially connected architecture are considered. Then the formulated optimization problem is given. A hybrid precoding and combining algorithm based on GSO is proposed in Section 3.3. In Section 3.4, computation complexity analysis is provided. The simulation results used to evaluate the effective and reliable performance of the proposed algorithm are provided in Section 3.5, which followed by a summary in Section 3.6.

3.2 System Model

Consider the single-user downlink communication of a fully connected and partially connected mmWave massive MIMO system based on hybrid precoding and combining as shown in Figure 3.1 and Figure 3.2, respectively. The BS with N_{t} antennas

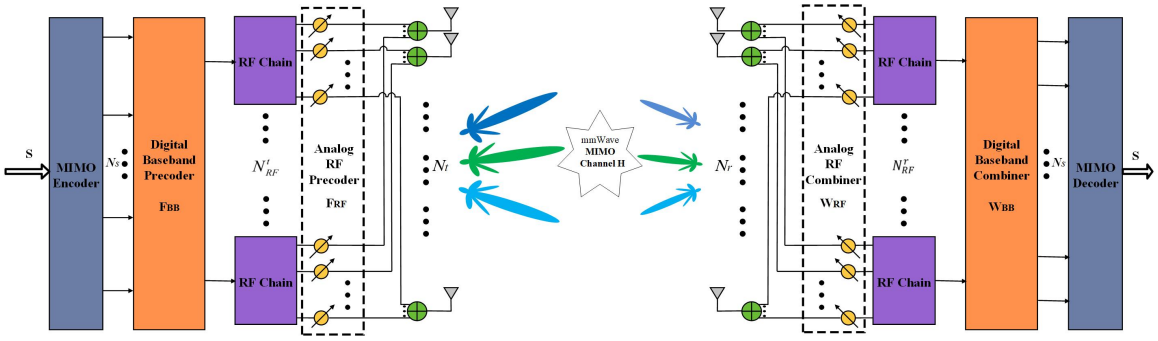

Figure 3.1: System block diagram of a fully connected mmWave hybrid precoding structure.

transmits N_{s} independent data streams to the MS with N_{r} antennas. Furthermore, we assume BS and MS are driven by a far smaller number of RF chains, namely, $N_{\mathrm{RF}}^{\mathrm{t}}$ and $N_{\mathrm{RF}}^{\mathrm{r}}$, respectively. To guarantee the effectiveness of the hybrid processing structure, the number of RF chains is constrained by $N_{\mathrm{s}} \leq N_{\mathrm{RF}}^{\mathrm{t}} \leq N_{\mathrm{t}}$ for the BS and $N_{\mathrm{s}} \leq N_{\mathrm{RF}}^{\mathrm{r}} \leq N_{\mathrm{r}}$ for the MS. This hybrid architecture enables the BS to apply an $N_{\mathrm{RF}}^{t} \times N_{\mathrm{s}}$ digital baseband precoder $\mathbf{F}_{\mathrm{BB}}$, followed by an $N_{\mathrm{t}} \times N_{\mathrm{RF}}^{\mathrm{t}}$ analog RF precoder $\mathbf{F}_{\mathrm{RF}}$, which is implemented through an array of analog phase shifters with constant norm. Therefore, all elements of $\mathbf{F}_{\mathrm{RF}}$ have equal norm. The only difference between Figure 3.1 and Figure 3.2 lies in that each RF chain is connected to all transmit antennas via phase shifters in Figure 3.1, while the antenna array at transmitter is divided into multiple sub-arrays and each sub-array is connected to multiple RF chains in Figure 3.2. In this case, each sub-array can transmit multiple data streams simultaneously. Furthermore, the sub-arrays are considered to be independent of each other in this chapter. The parallel structure is also applied to the receiver as well, i.e., the received signal is post-processed by an $N_{\mathrm{r}} \times N_{\mathrm{RF}}^{\mathrm{r}}$ analog RF combiner $\mathbf{W}_{\mathrm{RF}}$, followed by an $N_{\mathrm{RF}}^{\mathrm{r}} \times N_{\mathrm{s}}$ digital baseband combiner $\mathbf{W}_{\mathrm{BB}}$.

The spectral efficiency (bits/s/Hz) achieved by the mmWave hybrid precoding system when Gaussian symbols are transmitted over the channel with uniform power allocations among transmitted data streams, is given by

$$R = \log_2 \left(\left| \mathbf{I}_{N_{\mathrm{s}}} + \frac{P}{N_{\mathrm{s}}} \mathbf{R}_{\mathrm{n}}^{-1} \mathbf{G} \mathbf{G}^{H} \right| \right), \tag{3.1}$$

where P represents transmit power and $\mathbf{R}_{\mathrm{n}} = \sigma_{\mathrm{n}}^2 \mathbf{W}_{\mathrm{BB}}^{H} \mathbf{W}_{\mathrm{RF}}^{H} \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{BB}}$ denotes the covariance matrix of the combined noise. Each entry of the additive white noise vectors at N_{r} receive antennas follows the independent and identically distributed (i.i.d.) com-

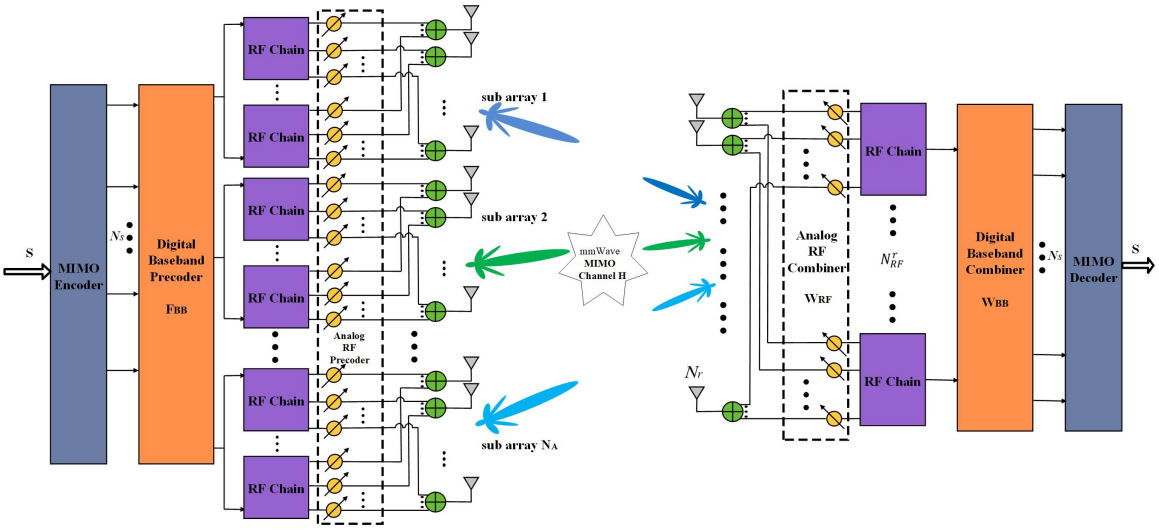

Figure 3.2: System block diagram of a partially connected mmWave hybrid precoding structure.

plex Gaussian distribution with zero mean and variance σ_{n}^2. $\mathbf{G} = \mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}$ denotes the equivalent processing matrix.

The optimal hybrid precoders $(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}})$ at BS and the optimal hybrid combiners $(\mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$ at MS can be obtained by maximizing the achieved spectral efficiency as given in Equation (3.1). Unfortunately, the non-convex constraints on $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ often makes finding the global optima intractable. A more feasible method is to decouple the joint transmitter-receiver optimization problem [58], and the decoupled hybrid precoder design at the transmitter end is formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}} \quad & \|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}\|_F \,, \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}}(:, i) \in \{\mathbf{a}_{\mathrm{BS}}(\phi_i), \forall i\} \,, \\ & \|\mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}\|_F^2 = N_s, \end{aligned} \tag{3.2}$$

where $\mathbf{F}_{\mathrm{opt}}$ denotes the optimal unconstrained precoder, which is composed of N_s right singular vectors corresponding to the N_s largest singular values of $\mathbf{H}$. The second constraint is the transmit power constraint. In this chapter, we only concentrate on the hybrid precoding design at the transmitter end, similarly, the hybrid combining design of $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ at the receiver end can be achieved in the same way.

The optimization problem illustrated by Equation (3.2) can be solved via the OMP algorithm [58]. The pseudo-code of OMP algorithm is summarized in Algorithm 1. In the OMP algorithm, when designing $\mathbf{F}_{\mathrm{BB}}$ via LS, matrix inversion is required by

N_{RF}^{t} times as shown in Step 7. To reduce calculation complexity, DFT matrix is adopted as a set of candidate RF precoder in OBMP algorithm [87]. However, the performance of OBMP will be decreased, and OBMP algorithm is constrained by array construction, which indicates that it is only applicable for ULA.

Algorithm 1 OMP Hybrid Precoding Algorithm

Input: $\mathbf{F}_{\mathrm{opt}}, \mathbf{A}_{\mathrm{t}}$
1. $\mathbf{F}_{\mathrm{RF}} = \mathrm{EmptyMatrix}$
2. $\mathbf{F}_{\mathrm{res}} = \mathbf{F}_{\mathrm{opt}}$
3. **for** $i \leq N_{\mathrm{RF}}^{\mathrm{t}}$ **do**
4. $\quad \psi = \mathbf{A}_{\mathrm{t}}^{H}\mathbf{F}_{\mathrm{opt}}$
5. $\quad k = \arg\max_{l=1,2,\cdots,Q}\left(\psi\psi^{H}\right)_{l,l}$
6. $\quad \mathbf{F}_{\mathrm{RF}} = [\mathbf{F}_{RF}|\,\mathbf{A}_{\mathrm{t}}^{k}]$
7. $\quad \mathbf{F}_{\mathrm{BB}} = \left(\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{opt}}$
8. $\quad \mathbf{F}_{\mathrm{res}} = \frac{\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}}{\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_{\mathrm{F}}}$
9. **end for**
10. $\mathbf{F}_{\mathrm{BB}} = \sqrt{N_{\mathrm{s}}}\frac{\mathbf{F}_{\mathrm{BB}}}{\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_{F}}$

Output $\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}$

3.3 Proposed Hybrid Precoding Algorithm

In this section, we present the proposed hybrid precoding algorithm. The rationale of the proposed algorithm is to utilize the columns of array response matrix orthogonalized by GSO as a set of candidate RF precoder, which can avoid the calculation of matrix inversion in designing digital precoding matrix. Since Gram Schmidt orthogonalization is a recursive process that depends on the order in which the matrix columns are selected. A heuristic on the order of columns selection should be according to the vector $\mathbf{a}_{\mathrm{BS}}(\varphi_i)$ along which the full digital precoder has the maximum projection. In other words, performing GSO to array response vectors obtained by the OMP algorithm. According to [97], given a set of N linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n$ from $\mathbb{C}^N$, we can construct an orthogonal set $\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_n$ which is linear combination of the original set and spans the same space as the one spanned by the original set.

The proposed algorithm utilizes the columns of array response matrix orthogonalized by GSO as a set of candidate RF precoder, then adopts OMP to find the optimal RF and BB precoder, which is named Prior GSO Orthogonal Matching Pur-

suit (PGOMP) algorithm. The PGOMP hybrid precoder design is formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}} \quad & \left\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_F, \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}}(:,i)\in\left\{\mathbf{A}_{\mathrm{t,orth}}(:,i),\forall i\right\}, \\ & {\mathbf{A}_{\mathrm{t,orth}}}^H(:,i)\mathbf{A}_{\mathrm{t,orth}}(:,j)=\begin{cases} 1, i=j \\ 0, i\neq j \end{cases} \\ & i,j=1,2,\cdots N_{\mathrm{c}}, \\ & \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_F^2=N_{\mathrm{s}}, \end{aligned} \tag{3.3}$$

where $\mathbf{A}_{\mathrm{t,orth}}$ denotes the unitary matrix obtained by performing GSO to the columns of array response matrix. The pseudo-code of PGOMP algorithm is given in Algorithm 2.

Algorithm 2 PGOMP Hybrid Precoding Algorithm

Input: $\mathbf{F}_{\mathrm{opt}}, \mathbf{A}_{\mathrm{t,orth}}$
Initialization: $\mathbf{U}_0 = \mathrm{EmptyMatrix}$
1. $\psi = \mathbf{A}_{\mathrm{t,orth}}^H \mathbf{F}_{\mathrm{opt}}$
2. $\mathbf{V} = \mathrm{diag}(\psi\psi^H)$
3. **for** $i \leq N_{\mathrm{RF}}^{\mathrm{t}}$ **do**
4. $\quad k = \arg\max_{1\leq j\leq N_C} \mathbf{V}(j)$
5. $\quad \mathbf{U}_i = [\mathbf{U}_{i-1} | k]$
6. $\quad \mathbf{V}(k) = 0$
7. **end for**
8. $\mathbf{F}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{t,orth}}(:, \mathbf{U})$
9. $\mathbf{F}_{\mathrm{BB}} = \psi(\mathbf{U}, :)$
10. $\mathbf{F}_{\mathrm{RF\_CM}} = \exp\left[\sqrt{-1} * \mathrm{phase}(\mathbf{F}_{\mathrm{RF}})\right]$
11. Calculate $\mathbf{\Lambda}$ according to (3.7).
12. $\mathbf{F}_{\mathrm{BB}} = \mathbf{\Lambda}\mathbf{F}_{\mathrm{BB}}$
13. $\mathbf{F}_{\mathrm{BB}} = \sqrt{N_{\mathrm{s}}} \frac{\mathbf{F}_{BB}}{\|\mathbf{F}_{\mathrm{RF\_CM}}\mathbf{F}_{\mathrm{BB}}\|_F}$
Output $\mathbf{F}_{\mathrm{RF\_CM}}, \mathbf{F}_{\mathrm{BB}}$

It can be seen from Algorithm 2, due to the orthogonality of $\mathbf{A}_{\mathrm{t,orth}}$, the calculation of matrix inversion can be avoided in designing $\mathbf{F}_{\mathrm{BB}}$ of step 9, as a result, the computation complexity can be decreased significantly. Furthermore, parallel process to find the best $N_{\mathrm{RF}}^{\mathrm{t}}$ vectors as $\mathbf{F}_{\mathrm{RF}}$ is utilized as described in OBMP Algorithm [87]. Moreover, according to [97], the space spanned by the columns of $\mathbf{A}_{\mathrm{t,orth}}$ is the same as the one spanned by the columns of array response matrix, as a result, the performance of PGOMP Algorithm is the same as the OMP Algorithm. However, it should be noted that the entries of $\mathbf{A}_{\mathrm{t,orth}}$ are not constant amplitude due to GSO, and hence the columns of $\mathbf{A}_{\mathrm{t,orth}}$ searched by OMP cannot be adopted as the RF

precoder directly. Therefore, the normalization step will be considered in the end of PGOMP algorithm, which may unfortunately cause some performance loss to some extent. Specifically, $\mathbf{F}_{\mathrm{RF}}$ is decoupled into the approximate product of two matrices, one being $\mathbf{F}_{\mathrm{RF\_CM}}$ with constant modulous entries, the other being a diagonal matrix M with dimension $N_{\mathrm{RF}}^{\mathrm{t}} \times N_{\mathrm{RF}}^{\mathrm{t}}$, i.e., $\mathbf{F}_{\mathrm{RF}} \approx \mathbf{F}_{\mathrm{RF\_CM}}\mathbf{\Lambda}$. The entries of $\mathbf{F}_{\mathrm{RF\_CM}}$ is composed of the phase of elements of $\mathbf{F}_{\mathrm{RF}}$, i.e., $\mathbf{F}_{\mathrm{RF\_CM}}(a,b) = e^{\sqrt{-1}*\mathrm{phase}(\mathbf{F}_{\mathrm{RF\_CM}}(a,b))}$, so $\mathbf{F}_{\mathrm{RF\_CM}}$ can be adopted as the RF precoder and the diagonal matrix $\mathbf{\Lambda}$ can be calculated according to Minimum Mean Square Error (MMSE) [99], which is given by

$$\begin{aligned} \min_{\mathbf{\Lambda}} \quad & \|\mathbf{F}_{\mathrm{RF}} - \mathbf{F}_{\mathrm{RF\_CM}}\mathbf{\Lambda}\|_F^2, \\ \text{s.t.} \quad & |\mathbf{F}_{\mathrm{RF\_CM}}(i,j)|^2 = 1, \forall i,j. \end{aligned} \tag{3.4}$$

It is obvious that the columns of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{RF\_CM}}$ are mutually independent. Therefore, the solving process in Equation (3.4) can be equivalent to the optimization problem for each column, i.e., minimizing the MMSE between $\mathbf{f}_{\mathrm{RF}}^i$ and $\mathbf{f}_{\mathrm{RF\_CM}}^i$. Now, the problem in Equation (3.4) can be simplified as

$$\begin{aligned} \min_{\lambda_i} \quad & \left\|\mathbf{f}_{\mathrm{RF}}^i - \mathbf{f}_{\mathrm{RF\_CM}}^i\lambda_i\right\|_F^2, \\ \text{s.t.} \quad & \left|\mathbf{f}_{\mathrm{RF\_CM}}^i(i,j)\right|^2 = 1, \forall i,j, \end{aligned} \tag{3.5}$$

where $\mathbf{f}_{\mathrm{RF}}^i$, $\mathbf{f}_{\mathrm{RF\_CM}}^i$ represent the ith column of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{RF\_CM}}$ respectively, λ_i is the ith element in the diagonal of $\mathbf{\Lambda}$. The Mean Square Error (MSE) can be further represented as

$$\begin{aligned} \|\mathbf{f}_{\mathrm{RF}}^i - \mathbf{f}_{\mathrm{RF\_CM}}^i\lambda_i\|_F^2 &= Tr\left(\mathbf{f}_{\mathrm{RF}}^i(\mathbf{f}_{\mathrm{RF}}^i)^H\right) - 2\lambda_i R\left((\mathbf{f}_{RF}^i)^H\mathbf{f}_{\mathrm{RF\_CM}}^i\right) + \lambda_i^2 Tr(\mathbf{f}_{\mathrm{RF\_CM}}^i{\mathbf{f}_{\mathrm{RF\_CM}}^i}^H) \\ &= \|\mathbf{f}_{\mathrm{RF}}^i\|_F^2 + \lambda_i^2\|\mathbf{f}_{\mathrm{RF\_CM}}^i\|_F^2 - 2\lambda_i R\left((\mathbf{f}_{\mathrm{RF}}^i)^H\mathbf{f}_{\mathrm{RF\_CM}}^i\right). \end{aligned} \tag{3.6}$$

Compute the partial derivatives of Equation (3.6), λ_i can be obtained by

$$\begin{aligned} \frac{\partial\left\|\mathbf{f}_{\mathrm{RF}}^i - \mathbf{f}_{\mathrm{RF\_CM}}^i\lambda_i\right\|_F^2}{\partial\lambda_i} &= 2\lambda_i\|\mathbf{f}_{\mathrm{RF\_CM}}^i\|_F^2 - 2(\mathbf{f}_{\mathrm{RF}}^i)^H\mathbf{f}_{\mathrm{RF\_CM}}^i = 0 \\ &\Rightarrow \lambda_i = \frac{(\mathbf{f}_{\mathrm{RF}}^i)^H\mathbf{f}_{\mathrm{RF\_CM}}^i}{\left\|\mathbf{f}_{\mathrm{RF\_CM}}^i\right\|_F^2}. \end{aligned} \tag{3.7}$$

Hence, the equivalent baseband precoder is $\mathbf{\Lambda}\mathbf{F}_{\mathrm{BB}}$, as illustrated from step 10 to 12 in Algorithm 2.

Then a heuristic solution on the order of columns selection is suggested to perform

GSO to $N_{\mathrm{RF}}^{\mathrm{t}}$ BS antenna response vectors along which the full digital precoder has the maximum the projection, which is named Orthogonal Matching Pursuit GSO (OMPG) algorithm. The OMPG hybrid precoder design is formulated as

$$\begin{aligned} \min_{\mathbf{F}_{\mathrm{RF}},\mathbf{F}_{\mathrm{BB}}} \quad & \left\|\mathbf{F}_{\mathrm{opt}}-\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_F, \\ \text{s.t.} \quad & \mathbf{F}_{\mathrm{RF}}(:,i)\in\{\mathbf{a}_{\mathrm{BS}}(\phi_i),\forall i\}, \\ & \text{GSO}: \ {\mathbf{F}_{\mathrm{RF}}}^H(:,i)\mathbf{F}_{\mathrm{RF}}(:,j)=\begin{cases} 1, i=j \\ 0, i\neq j \end{cases} \\ & i,j=1,2,\cdots N_{\mathrm{RF}}, \\ & \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right\|_F^2=N_{\mathrm{s}}. \end{aligned} \tag{3.8}$$

The pseudo-code of OMPG algorithm is shown in Algorithm 3. Via performing GSO to $N_{\mathrm{RF}}^{\mathrm{t}}$ BS antenna response vectors obtained by the OMP algorithm, the calculation of matrix inversion can also be avoided when designing $\mathbf{F}_{\mathrm{BB}}$, which can be seen from step 10 in Algorithm 3. Furthermore, the baseband precoder computational complexity of the OMPG algorithm can be largely reduced compared to that of Algorithm 2 (which requires performing GSO to N_{c} BS antenna response vectors in Algorithm 2, $N_{\mathrm{RF}}^{\mathrm{t}}<N_{\mathrm{c}}$). The constant norm of BS array response vectors becomes invalid due to GSO that is similar to Algorithm 2, and the same procedure (steps 13-15) should be followed in Algorithm 3.

After hybrid precoders $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ are obtained, we seek to design hybrid combiners $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$. In the absence of any hardware limitations, the exact optimal solution to combiner design is unconstrained MMSE combiner (full digital combiner) $\mathbf{W}_{\mathrm{MMSE}}$ [58], which is given by

$$\mathbf{W}_{\mathrm{MMSE}}=\frac{\sqrt{P}}{N_{\mathrm{s}}}\mathbf{F}_{\mathrm{BB}}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}^H\left(\frac{P}{N_{\mathrm{s}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^H\mathbf{F}_{\mathrm{RF}}^H\mathbf{H}^H+\sigma_n^2\mathbf{I}_{N_{\mathrm{r}}}\right)^{-1}. \tag{3.9}$$

The hybrid combiner $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ at the receiver can be obtained the same way, i.e., utilizing the columns of MS array response matrix orthogonalized by GSO as a set of candidate RF combiner and a heuristic solution to the order of columns selection is to perform GSO to $N_{\mathrm{RF}}^{\mathrm{r}}$ MS antenna response vectors along which full digital combiner $\mathbf{W}_{\mathrm{MMSE}}$ has the maximum projection. The pseudo-code of hybrid combiner corresponding to PGOMP is summarized in Algorithm 4. As mentioned before, the RF combiner matrix $\mathbf{W}_{\mathrm{RF}}$ also requires constant norm process, which can

Algorithm 3 OMPG Hybrid Precoding Algorithm

Input: $\mathbf{F}_{\text{opt}}, \mathbf{A}_{\text{t}}$
Initialization: $\mathbf{F}_{\text{RF}} = \text{EmptyMatrix}, \mathbf{F}_{\text{res}} = \mathbf{F}_{\text{opt}}$
1. **for** $i \leq N_{\text{RF}}^{\text{t}}$ **do**
2. $\psi = \mathbf{A}_{\text{t}}^{H}\mathbf{F}_{\text{res}}$
3. $k = \underset{1\leq l\leq N_C}{\arg\max} \left(\psi\psi^{H}\right)_{l,l}$
4. **if** $(i = 1)$
5. $\mathbf{F}_{\text{RF}}(:,1) = \mathbf{A}_{\text{t}}(:,k)$
6. **else**
7. $\mathbf{g}_{\text{RF}} = \mathbf{A}_{\text{t}}(:,k) - \sum_{j=1}^{i-1} \langle \mathbf{A}_{\text{t}}(:,k), \mathbf{F}_{\text{RF}}(:,j)\rangle \mathbf{F}_{\text{RF}}(:,j)$
8. $\mathbf{F}_{\text{RF}}(:,i) = \mathbf{g}_{\text{RF}} / \left\|\mathbf{g}_{\text{RF}}\right\|$
9. **end if**
10. $\mathbf{F}_{\text{BB}} = \mathbf{F}_{\text{RF}}^{H}\mathbf{F}_{\text{opt}}$
11. $\mathbf{F}_{\text{res}} = \frac{\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}\right\|_{\text{F}}}$
12. **end for**
13. $\mathbf{F}_{\text{RF\_CM}} = \exp\left[\sqrt{-1} * \text{phase}(\mathbf{F}_{\text{RF}})\right]$
14. Calculate $\mathbf{\Lambda}$ according to (3.7).
15. $\mathbf{F}_{\text{BB}} = \mathbf{\Lambda}\mathbf{F}_{\text{BB}}$
16. $\mathbf{F}_{\text{BB}} = \sqrt{N_{\text{s}}} \frac{\mathbf{F}_{\text{BB}}}{\left\|\mathbf{F}_{\text{RF\_CM}}\mathbf{F}_{\text{BB}}\right\|_{F}}$
Output $\mathbf{F}_{\text{RF\_CM}}, \mathbf{F}_{\text{BB}}$

been seen from step 10 to 12 in Algorithm 4.

The pseudo-code of hybrid combiner corresponding to OMPG is summarized in Algorithm 5. It can been seen from Algorithm 5, N_{RF}^{r} MS antenna response vectors obtained by OMP are performed GSO, which is shown from step 4 to 9. The RF combiner matrix $\mathbf{W}_{\text{RF}}$ also requires constant norm process, which can been seen from step 13 to 15.

3.4 Computational Complexity Analysis

The required computational complexity of PGOMP algorithm, OBMP algorithm, OMP algorithm and OMPG algorithm are presented and compared in Table 3.1. It can be observed from Table 3.1 that N_{c} and N_{RF}^{t} BS array response vectors are orthogonalized by GSO in the proposed PGOMP algorithm and OMPG algorithm, respectively. The additionally required computational complexity in the PGOMP algorithm and OMPG algorithm is $\mathcal{O}(N_{\text{c}} \times (2N_{\text{t}} + 1))$ and $\mathcal{O}(N_{\text{RF}}^{\text{t}} \times (2N_{\text{t}} + 1))$, respectively.

Furthermore, due to the orthogonalization characteristic of the RF precoder matrix in the PGOMP algorithm, OBMP algorithm and OMPG algorithm, the ma-

Algorithm 4 PGOMP Hybrid Combining Algorithm

Input: $\mathbf{W}_{\mathrm{MMSE}}, \mathbf{A}_{\mathrm{r,orth}}$
Initialization: $\mu_0 = \mathrm{EmptyMatrix}$
1. $\psi = \mathbf{A}_{\mathrm{r,orth.}}^{H} \mathbf{W}_{\mathrm{MMSE}}$
2.$\mathbf{V} = \mathrm{diag}(\psi\psi^{H})$
3. **for** $i \leq N_{\mathrm{RF}}^{\mathrm{r}}$ **do**
4. $\quad k = \underset{1 \leq l \leq N_C}{\arg\max} \left(\psi\psi^{H}\right)_{l,l}$
5. $\quad u_i = [u_{i-1} | \, k]$
6. $\quad \mathbf{V}(k) = 0$
7.**end for**
8. $\mathbf{W}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{r,orth}}(:, u)$
9. $\mathbf{W}_{\mathrm{BB}} = \psi(u, :)$
10. $\mathbf{W}_{\mathrm{RF\_CM}} = \exp\left[\sqrt{-1} * \mathrm{phase}(\mathbf{W}_{\mathrm{RF}})\right]$
11. Calculate $\mathbf{\Lambda}$ according to (3.7).
12. $\mathbf{W}_{\mathrm{BB}} = \mathbf{\Lambda}\mathbf{W}_{\mathrm{BB}}$
13. $\mathbf{W}_{\mathrm{res}} = \frac{\mathbf{W}_{\mathrm{MMSE}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}}{\|\mathbf{W}_{\mathrm{MMSE}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\|_F}$
Output $\mathbf{W}_{\mathrm{RF\_CM}}, \mathbf{W}_{\mathrm{BB}}$

Algorithm 5 OMPG Hybrid Combining Algorithm

Input: $\mathbf{W}_{\mathrm{MMSE}}, \mathbf{A}_{\mathrm{r}}$
Initialization: $\mathbf{W}_{\mathrm{res}} = \mathbf{W}_{\mathrm{MMSE}}, \mathbf{W}_{\mathrm{RF}} = \mathrm{EmptyMatrix}$
1. **for** $i \leq N_{\mathrm{RF}}^{\mathrm{r}}$ **do**
2. $\quad \psi = \mathbf{A}_{\mathrm{r}}^{H} \mathbf{W}_{\mathrm{MMSE}}$
3. $\quad k = \underset{1 \leq l \leq N_{\mathrm{C}}}{\arg\max} \left(\psi\psi^{H}\right)_{l,l}$
4. $\quad$ **if** $(i = 1)$
5. $\quad\quad \mathbf{W}_{\mathrm{RF}}(:, 1) = \mathbf{A}_{\mathrm{r}}(:, k)$
6. $\quad$ **else**
7. $\quad\quad \mathbf{g}_{\mathrm{RF}} = \mathbf{A}_{\mathrm{r}}(:, k) - \sum_{j=1}^{i-1} \langle \mathbf{A}_{\mathrm{r}}(:, k), \mathbf{W}_{\mathrm{RF}}(:, j) \rangle \mathbf{W}_{\mathrm{RF}}(:, j)$
8. $\quad\quad \mathbf{W}_{\mathrm{RF}}(:, i) = \mathbf{g}_{\mathrm{RF}} / \|\mathbf{g}_{\mathrm{RF}}\|$
9. $\quad$ **end if**
10. $\quad \mathbf{W}_{\mathrm{BB}} = \mathbf{W}_{\mathrm{RF}}^{H} \mathbf{W}_{\mathrm{MMSE}}$
11. $\quad \mathbf{W}_{\mathrm{res}} = \frac{\mathbf{W}_{\mathrm{MMSE}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}}{\|\mathbf{W}_{\mathrm{MMSE}} - \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\|_{\mathrm{F}}}$
12. **end for**
13. $\mathbf{W}_{\mathrm{RF\_CM}} = \exp\left[\sqrt{-1} * \mathrm{phase}(\mathbf{W}_{\mathrm{RF}})\right]$
14. Calculate $\mathbf{\Lambda}$ according to (3.7).
15. $\mathbf{W}_{\mathrm{BB}} = \mathbf{\Lambda}\mathbf{W}_{\mathrm{BB}}$
Output $\mathbf{W}_{\mathrm{RF\_CM}}, \mathbf{W}_{\mathrm{BB}}$

trix inversion can be avoided when designing the digital precoding matrix $\mathbf{F}_{\mathrm{BB}}$, which also means, $\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{opt}}$. Then, the required computational complexity is $\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$. However, matrix inversion is required in designing the digital precoder using the OMP algorithm, i.e., $\mathbf{F}_{\mathrm{BB}} = \left(\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{F}_{\mathrm{RF}}^{H}\mathbf{F}_{\mathrm{opt}}$, and the required computational complexity is $\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} (k^2 + kN_{\mathrm{s}})N_{\mathrm{t}} + k^3)$. Therefore, the additionally required computational complexity in the PGOMP algorithm is $\mathcal{O}(N_{\mathrm{c}} \times (2N_{\mathrm{t}} + 1))$ compared to the OBMP algorithm. And the additionally required computational complexity in the OMP algorithm is $\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} (k^2N_{\mathrm{t}} + k^3) - N_{\mathrm{RF}}^{t}(2N_{\mathrm{t}} + 1))$ compared to the OMPG algorithm. Compared to the PGOMP algorithm, the additionally required computational complexity in the OMPG algorithm is $\mathcal{O}((N_{\mathrm{RF}}^{\mathrm{t}} - 1)N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + (N_{\mathrm{RF}}^{\mathrm{t}} - N_{\mathrm{c}})(2N_{\mathrm{t}} + 1))$.

Table 3.1: Comparison of computational complexity

Algorithms	**Calculate $\mathbf{F_{RF}}$**	**Calculate $\mathbf{F_{BB}}$**	**Total Complexity**
PGOMP	$\mathcal{O}(N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + N_{\mathrm{c}}(2N_{\mathrm{t}} + 1))$	$\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$	$\mathcal{O}(N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + N_{\mathrm{c}}(2N_{\mathrm{t}} + 1) + \sum_{k=1}^{N_{RF}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$
OBMP [27]	$\mathcal{O}(N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}))$	$\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$	$\mathcal{O}(N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + \sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$
OMP [22]	$\mathcal{O}(N_{\mathrm{RF}}^{\mathrm{t}}N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}))$	$\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} (k^2N_{\mathrm{t}} + kN_{\mathrm{s}}N_{\mathrm{t}} + k^3))$	$\mathcal{O}(N_{\mathrm{RF}}^{\mathrm{t}}N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + \sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} (k^2N_{\mathrm{t}} + kN_{\mathrm{s}}N_{\mathrm{t}} + k^3))$
OMPG	$\mathcal{O}(N_{\mathrm{RF}}^{\mathrm{t}}N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + N_{\mathrm{RF}}^{\mathrm{t}}(2N_{\mathrm{t}} + 1))$	$\mathcal{O}(\sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$	$\mathcal{O}(N_{\mathrm{RF}}^{\mathrm{t}}N_{\mathrm{c}}N_{\mathrm{s}}(N_{\mathrm{t}} + N_{\mathrm{c}}) + N_{\mathrm{RF}}^{\mathrm{t}}(2N_{\mathrm{t}} + 1) + \sum_{k=1}^{N_{\mathrm{RF}}^{\mathrm{t}}} kN_{\mathrm{s}}N_{\mathrm{t}})$

Take $N_{\mathrm{t}} = 256$, $N_{\mathrm{c}} = 8$, $N_{\mathrm{RF}}^{\mathrm{t}} = 6$ and $N_{\mathrm{s}} = 6$ as an example. The computational complexity of the RF precoder, baseband precoder, and total precoder is shown in Figure 3.3. The four groups are PGOMP, OBMP, OMP, and OMPG algorithms, respectively. It can be seen from Figure 3.3 that the computational complexity of the proposed PGOMP and OMPG algorithms significantly decreased compared to the OMP algorithm.

3.5 Simulation Results

In this section, simulation results are presented to illustrate the SE and BER performance of the proposed algorithms. We consider an environment with 10 scatters, and after performing QPSK modulation, the length of each data stream is 100 symbols. For the full digital precoding algorithm, the precoding matrix is composed of

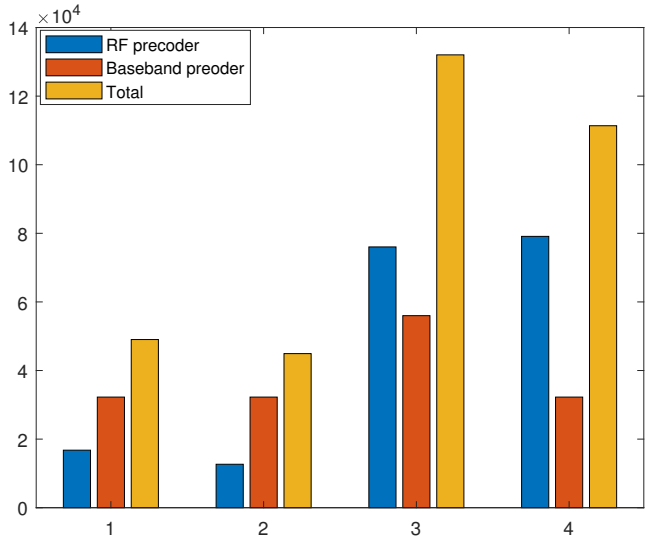

Figure 3.3: Comparison of computational complexity with fully connected structure, $N_{\mathrm{t}} = 256$, $N_{\mathrm{c}} = 8$, $N_{\mathrm{RF}}^{\mathrm{t}} = 6$, and $N_{\mathrm{s}} = 6$.

N_s right singular vectors corresponding to the N_s largest singular values of $\mathbf{H}$ and the combining matrix is composed of N_s left singular vectors corresponding to the N_s largest singular values of $\mathbf{H}$. In order to implement full digital precoding and combining algorithm, a dedicated RF chain is required for each antenna element both at the transmitter and the receiver. Furthermore, it requires perfect CSI to perform singular value decomposition of $\mathbf{H}$. For the rest algorithms, the full digital MMSE combiner is adopted at the receiver, and all the results are simulated over 3000 channel realizations. The signal-to-noise ratio is defined as SNR $= P/\sigma_n^2$. Figure 3.4 to Figure 3.9 show the performance comparison with fully connected hybrid precoding mmWave massive MIMO systems assuming perfect CSI, while Figure 3.10 to Figure 3.17 show performance comparison with partially connected (PC) hybrid precoding mmWave massive MIMO systems assuming perfect CSI. Figure 3.18 to Figure 3.21 show the performance of the proposed algorithms with imperfect CSI. Noted that for PC scenarios, only one RF chain is assumed to be connected to each predetermined subarray in our simulation. In this case, it should be noted since the RF precoding of each subarray is a vector, the proposed PGOMP algorithm is more suitable for the PC hybrid precoding mmWave massive MIMO systems. It is straightforward to extend one RF chain to multiple RF chains.

Figure 3.4 and Figure 3.5 show the SE and BER performance comparison with

ULA at both BS and MS end, with $N_{\mathrm{t}} = 64$, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ RF chains. It can be seen from Figure 3.4 that the SE performance of the proposed PGOMP and OMPG algorithm are almost the same as the OMP algorithm and full digital precoding in the case $N_{\mathrm{s}} = 2$ and are almost the same as the OMP algorithm and within a small gap from full digital precoding algorithm in the case $N_{\mathrm{s}} = 4$ and 6 data streams. It should be noted the proposed OMPG algorithm outperforms the proposed PGOMP algorithm which confirms our heuristic assumption on the order of columns selection to perform GSO. Further, we note that although OBMP algorithm can also avoid matrix inversion when designing $\mathbf{F}_{\mathrm{BB}}$, the SE performance gain of the proposed algorithm (with SNR larger than -10 dB) compared to OBMP algorithm when transmitting 2, 4 and 6 streams, is approximately 1%, 3% and 6%, respectively. Figure 3.5 shows the proposed algorithms achieve almost perfect BER performance in the case when $N_{\mathrm{s}} = 2$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{s}} = 6$. Further, we note the BER performance gain of the proposed algorithm (with BER equal to 10^{-2}) compared to OBMP algorithm when transmitting 2, 4 and 6 streams, is approximately 0.6 dB, 1.0 dB and 2.0 dB, respectively.

Figure 3.6 and Figure 3.7 plot the SE and BER performance comparison with ULA at both BS and MS end, with $N_{\mathrm{t}} = 256$, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ RF chains. It can be seen from Figure 3.6 that the SE performance gain of the proposed algorithm (with SNR larger than -10 dB) compared to OBMP algorithm when transmitting 2, 4 and 6 streams is approximately 1% , 3% and 7%, respectively. Figure 3.7 shows the BER performance gain of the proposed algorithm (with BER equal to 10^{-2}) compared to OBMP algorithm when transmitting 2, 4 and 6 streams, is approximately 0.4 dB, 0.5 dB and 6.0 dB, respectively.

To explore the performance when using UPA in mmWave systems, Figure 3.8 and Figure 3.9 plot the SE and BER performance comparison when using UPA at both BS and MS ends, with $N_{\mathrm{t}} = 16 \times 16$, $N_{\mathrm{r}} = 8 \times 8$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ RF chains. As illustrated in Figure 3.8 and Figure 3.9, considering the SE and BER performance comparison, both the proposed PGOMP and OMPG algorithms are not affected by array configuration. However, since the space spanned by the columns of DFT matrix is different from the one spanned by the array response vectors of UPA, the SE performance of OBMP Algorithm is largely outperformed by other algorithms particularly when SNR increases. In this case, two-dimensional DFT should be considered

Figure 3.10 and Figure 3.11 show the SE and BER performance comparison with

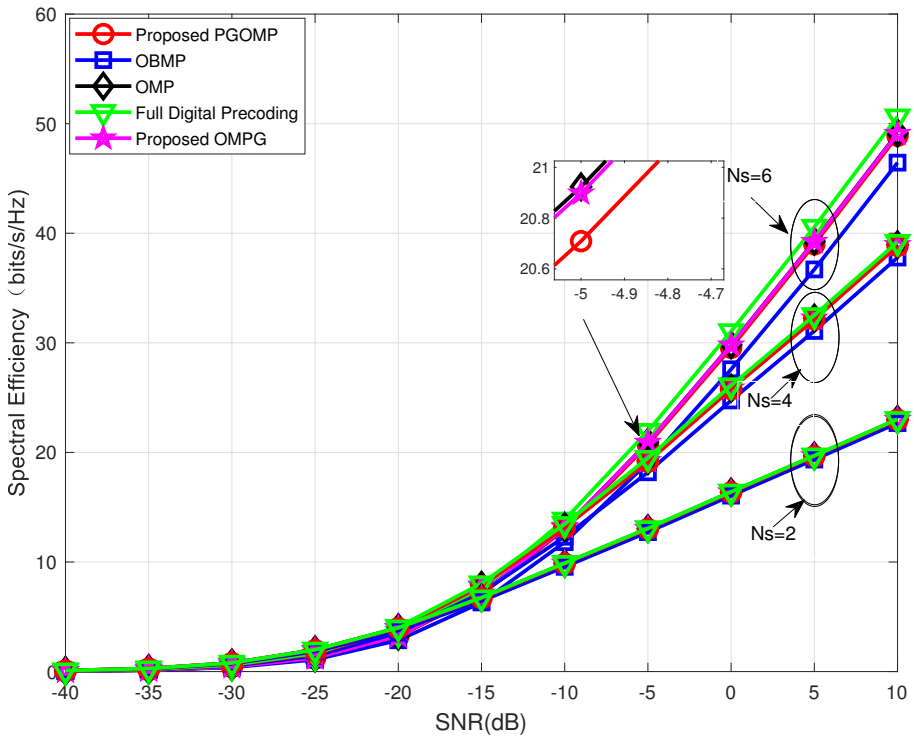

Figure 3.4: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64$, fully connected structure, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

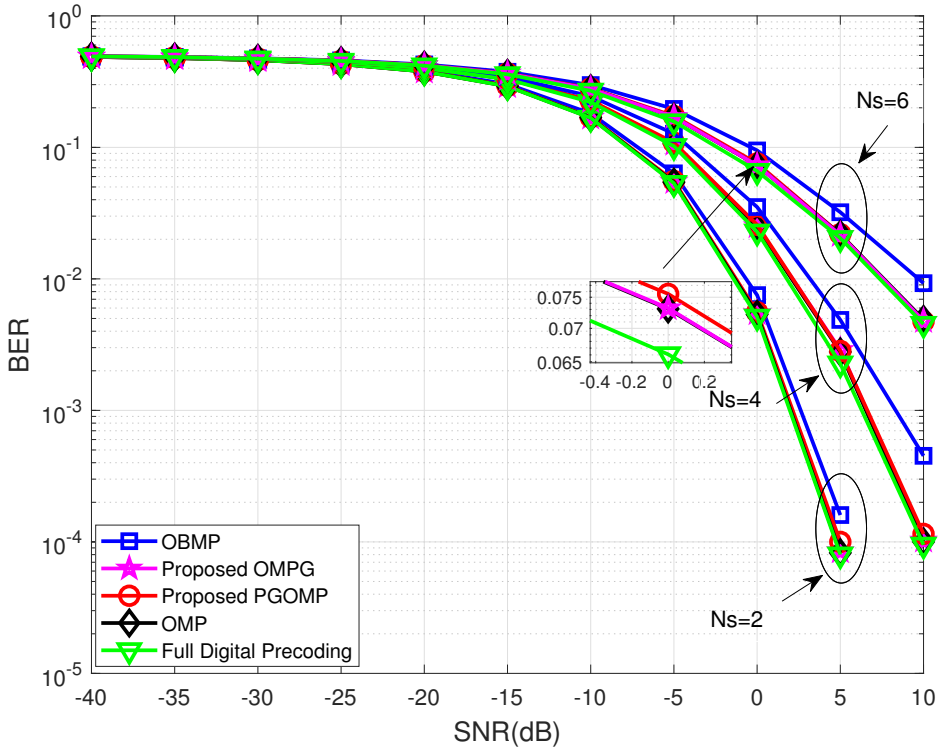

Figure 3.5: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64$, fully connected structure, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

ULA at both BS and MS end, with $N_{\mathrm{t}} = 64$ (the antenna number of each subarray is divided by N_{s}), $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$ RF chains. It can be seen from Figure 3.10 that the SE performance of the PC proposed PGOMP algorithm is almost the same as the OMP algorithm and within a small gap from full digital precoding algorithm in the case $N_{\mathrm{s}} = 2$ and $N_{\mathrm{s}} = 4$ data streams. And the SE performance gain of the proposed algorithm (with SNR larger than -10 dB) compared to OBMP algorithm when transmitting 2 and 4 streams, is approximately 4% and 2%, respectively. Figure 3.11 shows that the BER performance of the PC Proposed PGOMP algorithm is within a small gap from OMP algorithm in the cases when $N_{\mathrm{s}} = 2$ and $N_{\mathrm{s}} = 4$ data streams. And the BER performance gain of the proposed algorithm (with BER equal to 10^{-3}) compared to OBMP algorithm when transmitting 2, and 4 streams, is approximately 0.6 dB and 0.5 dB, respectively.

Figure 3.12 and Figure 3.13 show the SE and BER performance comparison with ULA at both BS and MS end, with $N_{\mathrm{t}} = 256$ (the antenna number of each subarray is divided by N_{s}), $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$ RF chains. Figure 3.12 shows that

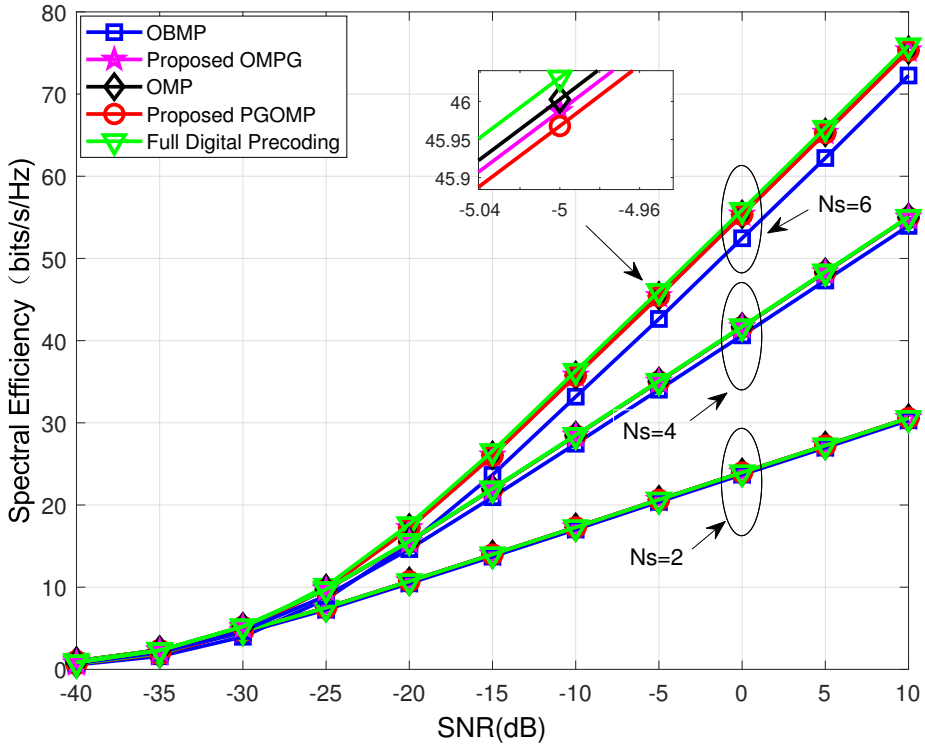

Figure 3.6: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 256$, fully connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

the SE performance of the PC Proposed PGOMP algorithm is almost the same as the OMP algorithm and within a small gap from full digital precoding algorithm in the cases when $N_{\mathrm{s}} = 2$ and $N_{\mathrm{s}} = 4$ data streams. And the SE performance gain of the proposed algorithm (with SNR larger than -10 dB) compared to OBMP algorithm when transmitting 2 and 4 streams, is approximately 4% and 2%, respectively. It can be observed from Figure 3.13 that the BER performance of the PC Proposed PGOMP algorithm is almost the same as the OMP algorithm in the cases when $N_{\mathrm{s}} = 2$ and $N_{\mathrm{s}} = 4$ data streams. And the BER performance gain of the proposed algorithm (with BER equal to 10^{-3}) compared to OBMP algorithm when transmitting 2, and 4 streams, is approximately 0.6 dB and 0.5 dB, respectively.

Figure 3.14 and Figure 3.15 show the SE and BER performance comparison with ULA at both BS and MS end, with the antenna number of each sub-array is 64, $N_{\mathrm{t}} = 64 \times N_{\mathrm{s}}$, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$ RF chains. It can be seen from Figure 3.14 that the SE performance of the PC Proposed PGOMP algorithm are almost the same as the OMP algorithm and within a small gap from full digital

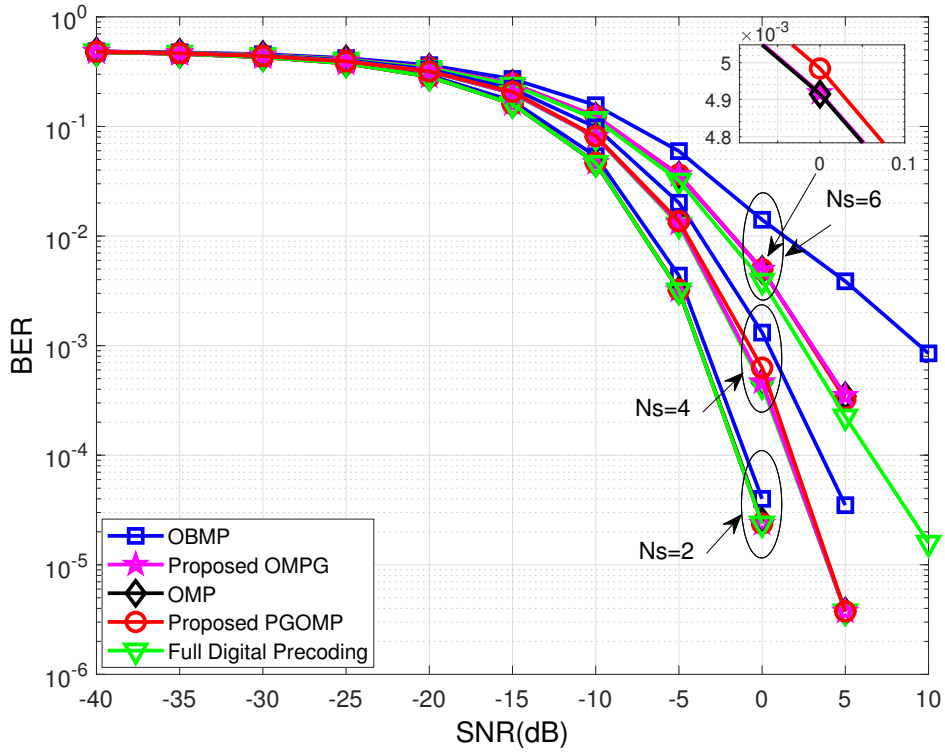

Figure 3.7: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 256$, fully connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

precoding algorithm in the case $N_{\mathrm{s}} = 2$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{s}} = 6$ data streams. And the SE performance gain of the proposed algorithm (with SNR larger than -10 dB) compared to OBMP algorithm when transmitting 2, 4 and 6 streams, is approximately 4%, 5% and 5%, respectively. Figure 3.15 shows that the BER performance of the PC Proposed PGOMP algorithm is almost the same as the OMP algorithm and within a small gap from full digital precoding algorithm in the cases when $N_{\mathrm{s}} = 2$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{s}} = 6$ data streams. And the BER performance gain of the proposed algorithm (with BER equal to 10^{-2}) compared to OBMP algorithm when transmitting 2, 4 and 6 streams, is approximately 0.7 dB, 0.8 dB and 1.0 dB, respectively.

To explore performance in partially connected mmWave systems with UPA, Figure 3.16 and Figure 3.17 plot the SE and BER performance comparison with UPA subarray at BS and UPA at MS end, with the antenna number of each subarray is 64, $N_{\mathrm{s}} = 6$, $N_{\mathrm{t}} = 64 \times N_{\mathrm{s}}$, $N_{\mathrm{r}} = 4 \times 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$ RF chains. As illustrated in Figure 3.16 and Figure 3.17, considering the SE and BER performance comparison with UPA at both BS and MS end, the PC proposed PGOMP algorithm are not

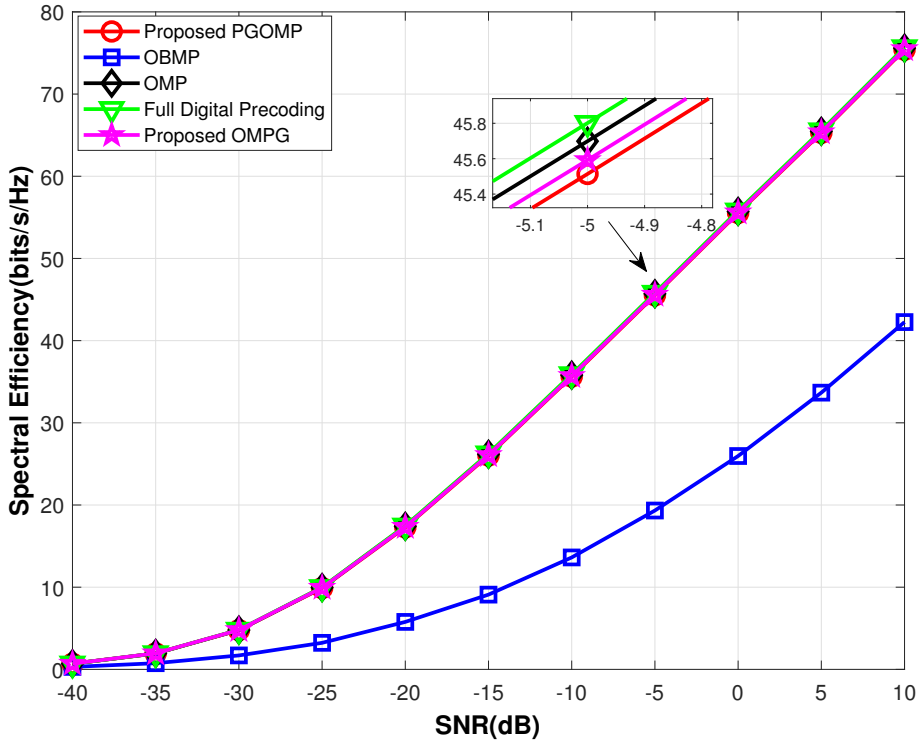

Figure 3.8: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 16 \times 16$, fully connected structure, $N_{\mathrm{r}} = 8 \times 8$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

affected by array configuration. However, since the space spanned by the columns of DFT matrix is different from the one spanned by the array response vectors of UPA, the SE performance of PC OBMP algorithm is largely outperformed by other algorithms particularly when SNR increases. In this case, two-dimensional DFT should be considered.

Furthermore, Figure 3.18 to Figure 3.21 evaluate the impact of imperfect CSI on the proposed algorithm. The estimated channel matrix can be modeled as [17]

$$\hat{\mathbf{H}} = \xi \mathbf{H} + \sqrt{1-\xi^2}\mathbf{E}, \tag{3.10}$$

where $0 \leq \xi \leq 1$ represents the reliability of the estimate and $\mathbf{E}$ is an error matrix with i.i.d. $\mathcal{CN}(0,1)$ distributed entries.

Figure 3.18 and Figure 3.19 show the SE and BER performance in fully connected structure with $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ and $N_{\mathrm{s}} = 4$. It can be seen from Figure 3.18 that the

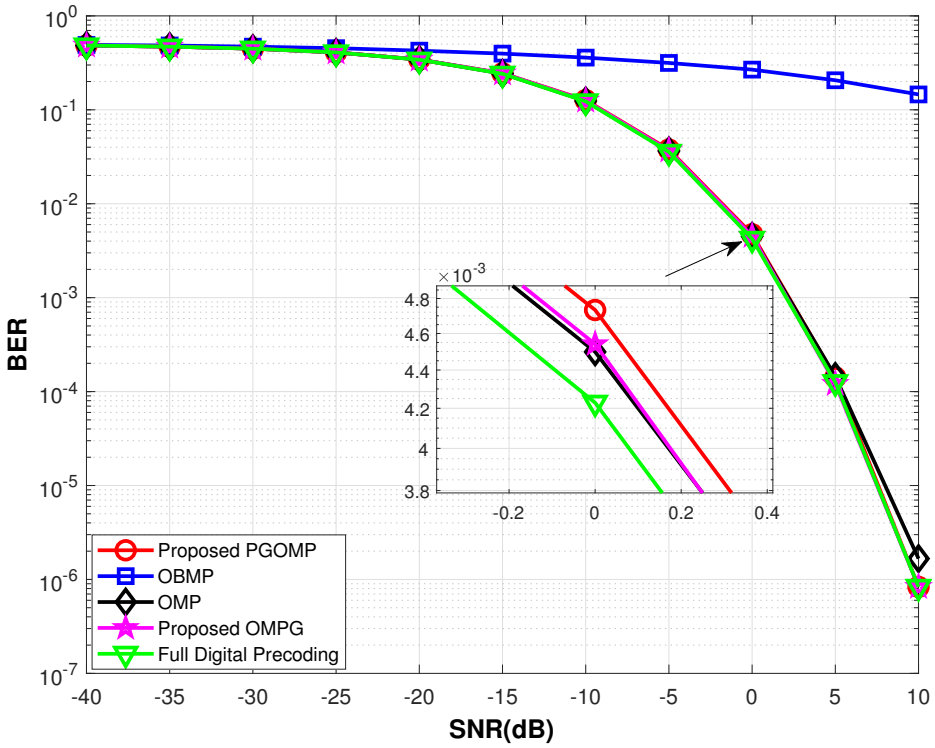

Figure 3.9: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 16 \times 16$, fully connected structure, $N_{\mathrm{r}} = 8 \times 8$ and $N_{\mathrm{RF}}t = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

SE performance loss of the Proposed OMPG and PGOMP algorithm with imperfect CSI ($\xi = 0.9$ and $\xi = 0.7$, Im CSI is used in the figure) compared to perfect CSI is approximately 1 bit/s/Hz and 4 bits/s/Hz, respectively. Figure 3.19 shows that the BER performance loss of the Proposed OMPG and PGOMP algorithm with imperfect CSI ($\xi = 0.9$ and $\xi = 0.7$, Im CSI is used in the figure) compared to perfect CSI is approximately 0.7 dB and 1.9 dB (with BER equal to 10^{-2}), respectively. It also can be seen from Figure 3.18 and Figure 3.19 that the SE and BER performance of the Proposed OMPG and PGOMP algorithm are almost the same.

Figure 3.20 and Figure 3.21 show the SE and BER performance in partially connected structure with $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 4$ and $N_{\mathrm{s}} = 4$. It can be seen from Figure 3.20 that the SE performance loss of the Proposed PGOMP algorithm with imperfect CSI ($\xi = 0.9$ and $\xi = 0.7$) compared to perfect CSI is approximately 1 bit/s/Hz and 4 bits/s/Hz, respectively. Figure 3.21 shows that the BER performance loss of the Proposed PGOMP algorithm with imperfect CSI ($\xi = 0.9$ and $\xi = 0.7$) compared to perfect CSI is approximately 0.7 dB and 3 dB (with BER equal to 10^{-2}), respectively.

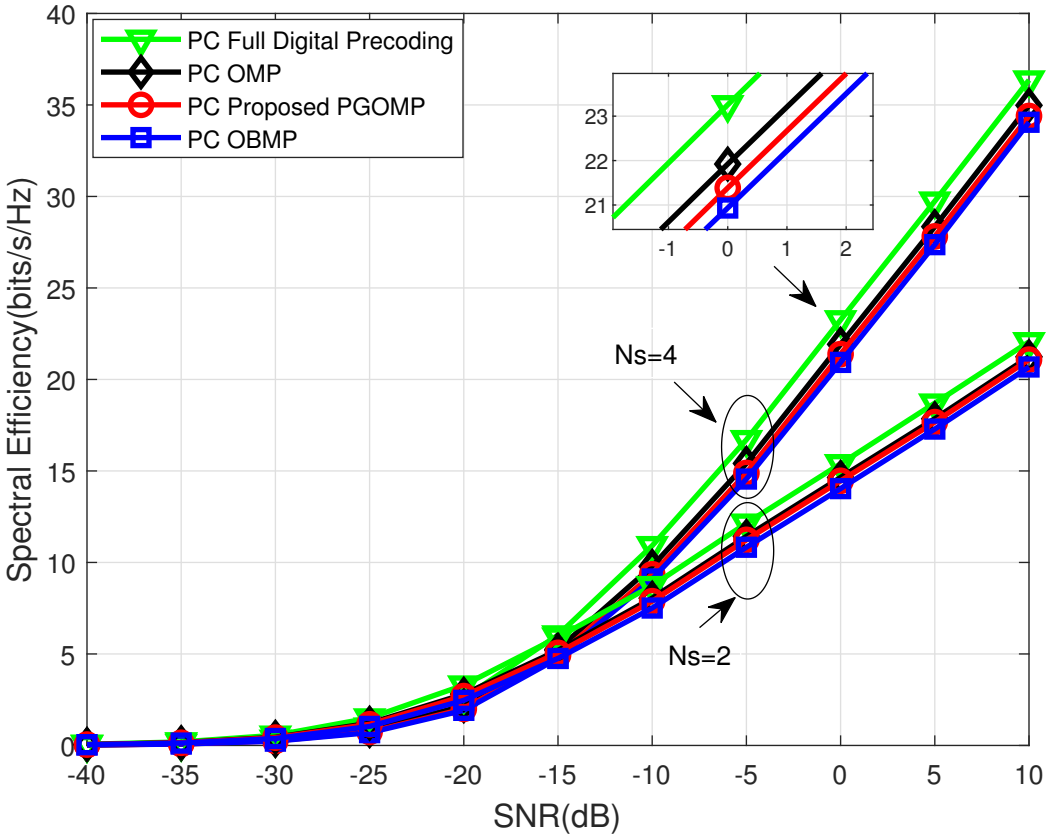

Figure 3.10: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64$, partially connected structure, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

From Figure 3.18 to Figure 3.21, we can observe that the proposed algorithm is able to cope with imperfect CSI.

3.6 Summary

In this chapter, we investigated single user GSO based hybrid precoding and combining algorithms in both fully and partially connected mmWave massive MIMO architectures. The computation complexity of the proposed algorithm is decreased significantly compared to the OMP algorithm, while it is increased a bit more than OBMP algorithm. Furthermore, the proposed algorithm is unconstrained to array construction. Finally, we presented simulation results on the performance of the proposed GSO based hybrid precoding and combining algorithm, the simulation results

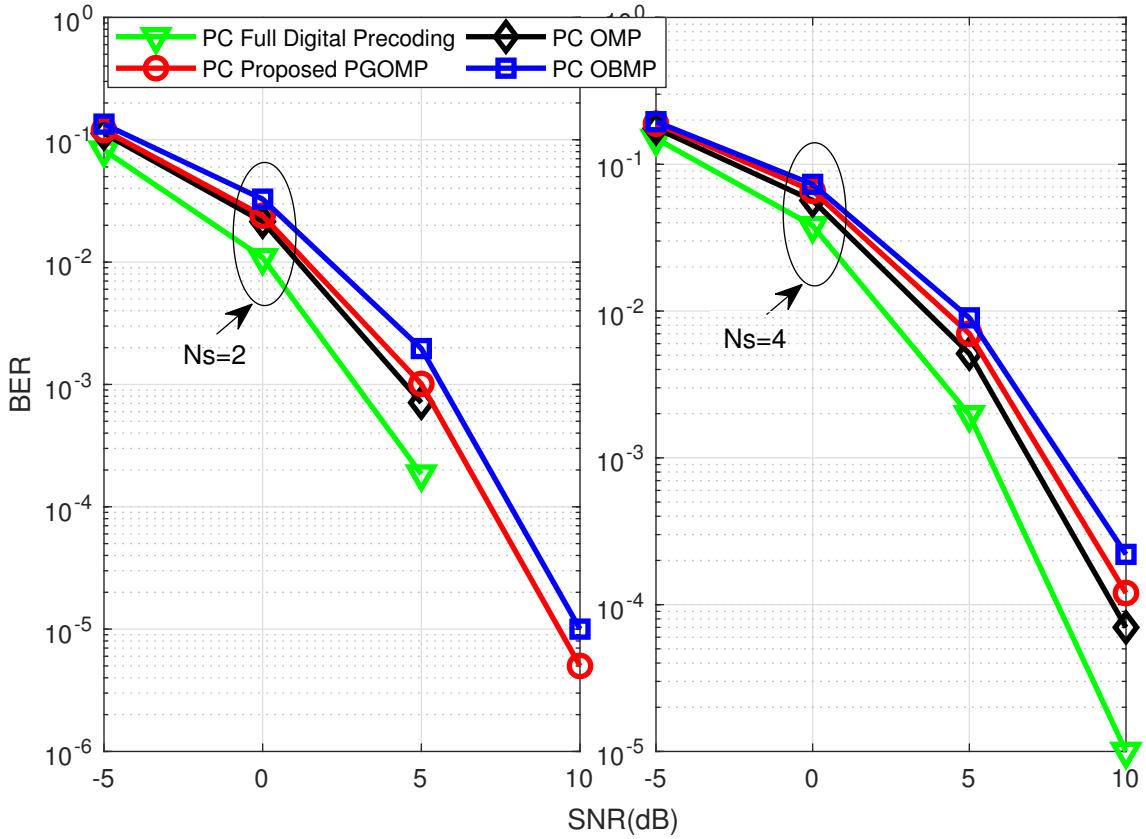

Figure 3.11: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64$, partially connected structure, $N_{\mathrm{r}} = 16$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

have verified that the proposed algorithm outperforms OBMP algorithm and is almost the same as the OMP algorithm on SE and BER performance.

It is worth mentioning that the proposed algorithm is based on the narrowband flat fading mmWave channel and designed for the single user scenario. In the next chapter, we will develop efficient hybrid precoding and combining algorithm for multi-user scenarios, which may emerge more frequently in future applications.

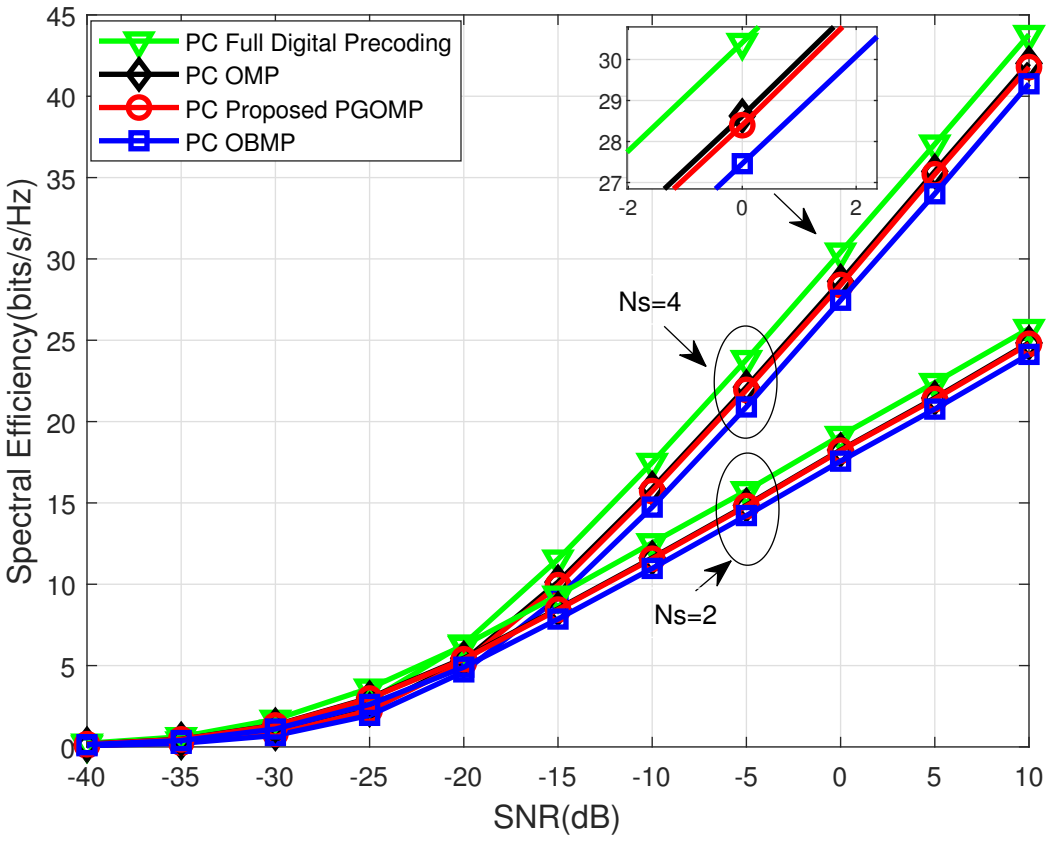

Figure 3.12: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 256$, partially connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

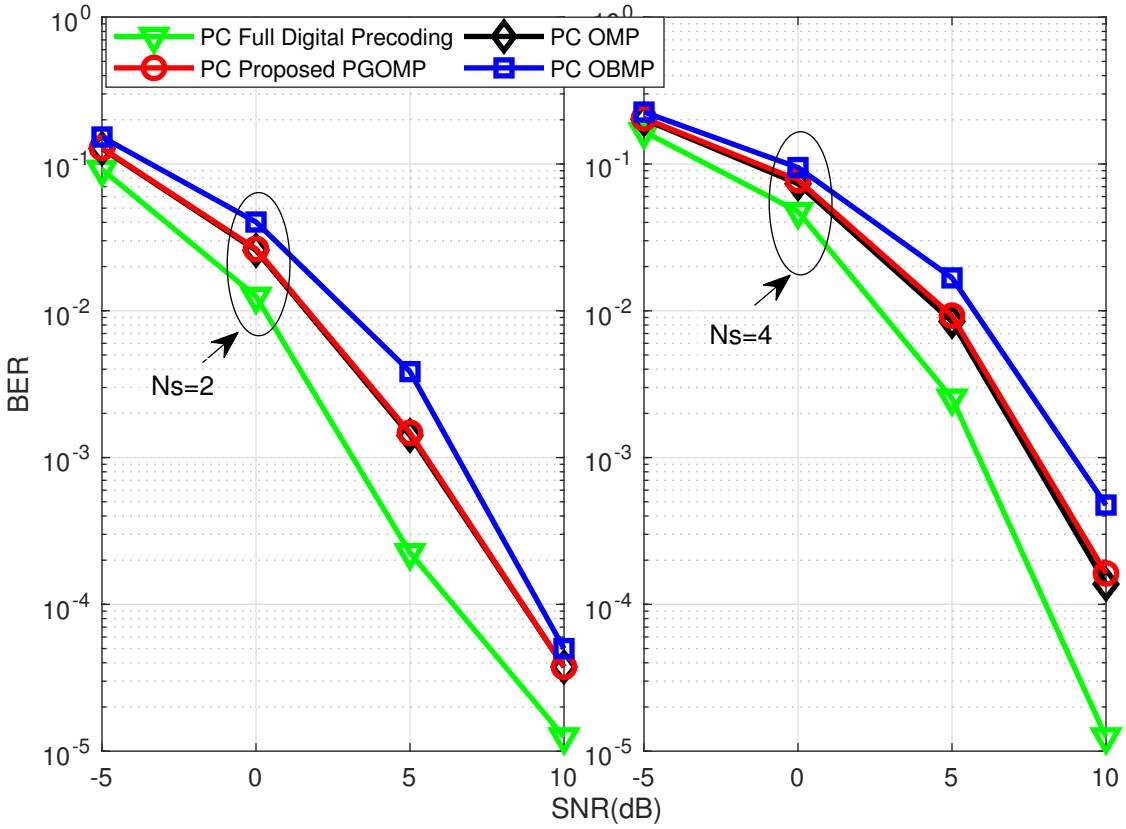

Figure 3.13: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 256$, partially connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

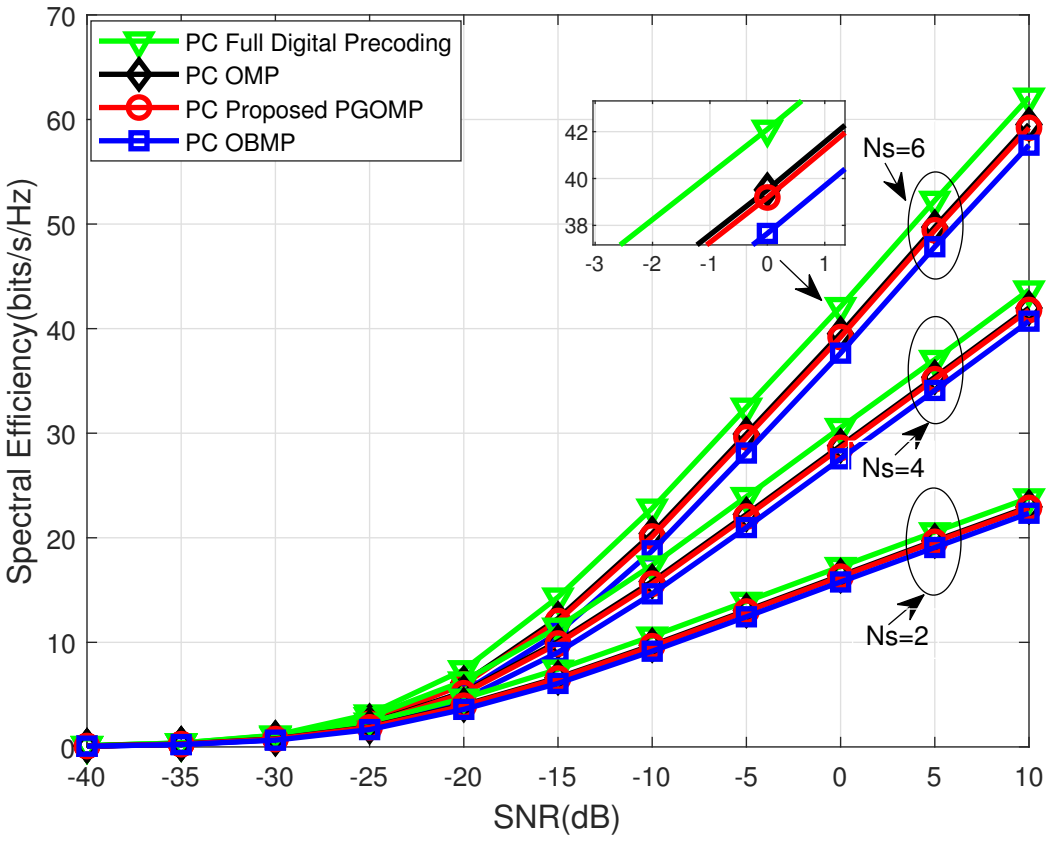

Figure 3.14: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64 \times N_{\mathrm{s}}$, partially connected structure, $N_{\mathrm{r}} = 16$, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

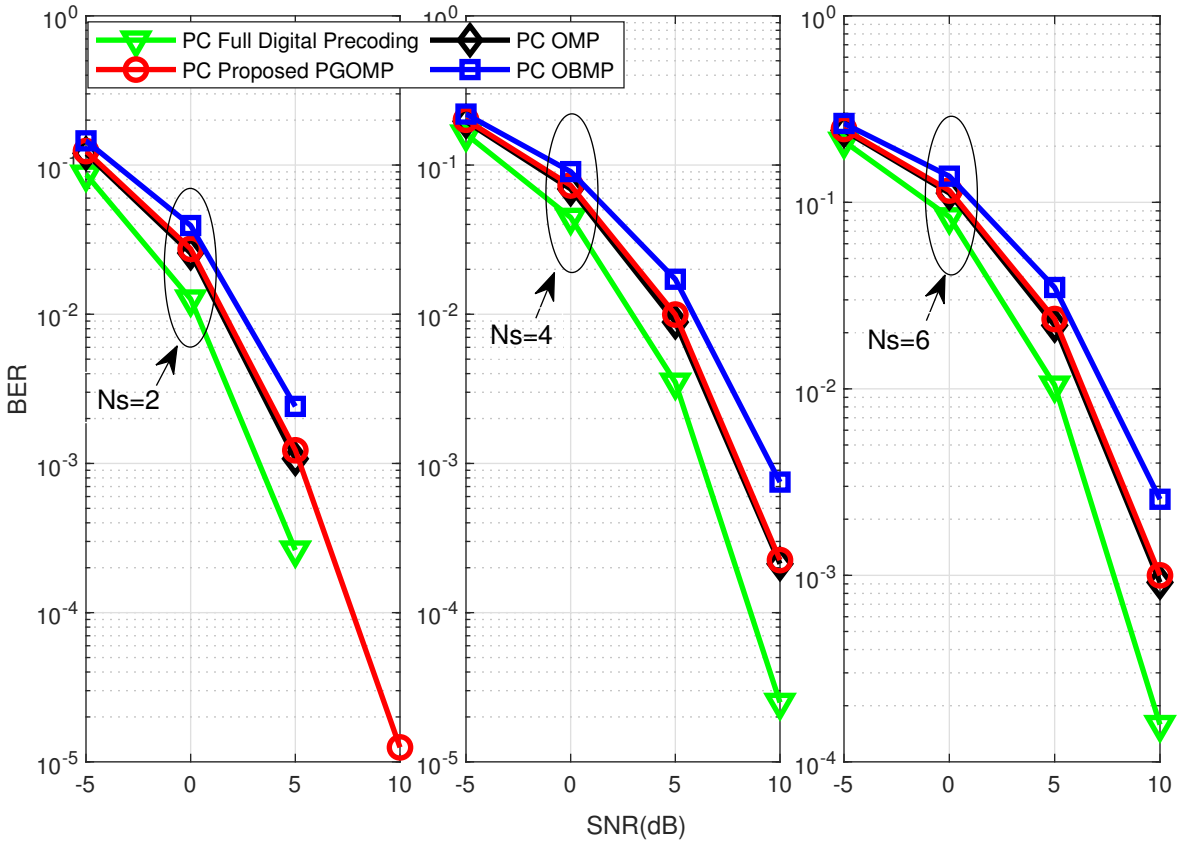

Figure 3.15: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = 64 \times N_{\mathrm{s}}$, partially connected structure, $N_{\mathrm{r}} = 16$, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$.

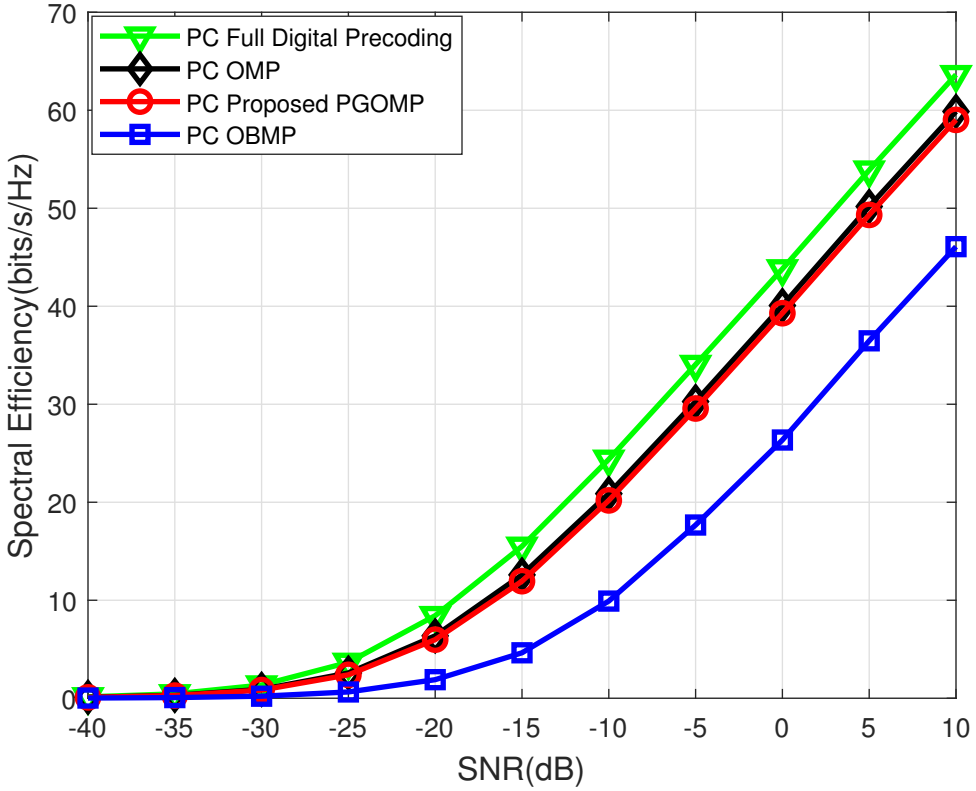

Figure 3.16: Spectral efficiency performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = (16 \times 16) \times N_{\mathrm{s}}$, partially connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

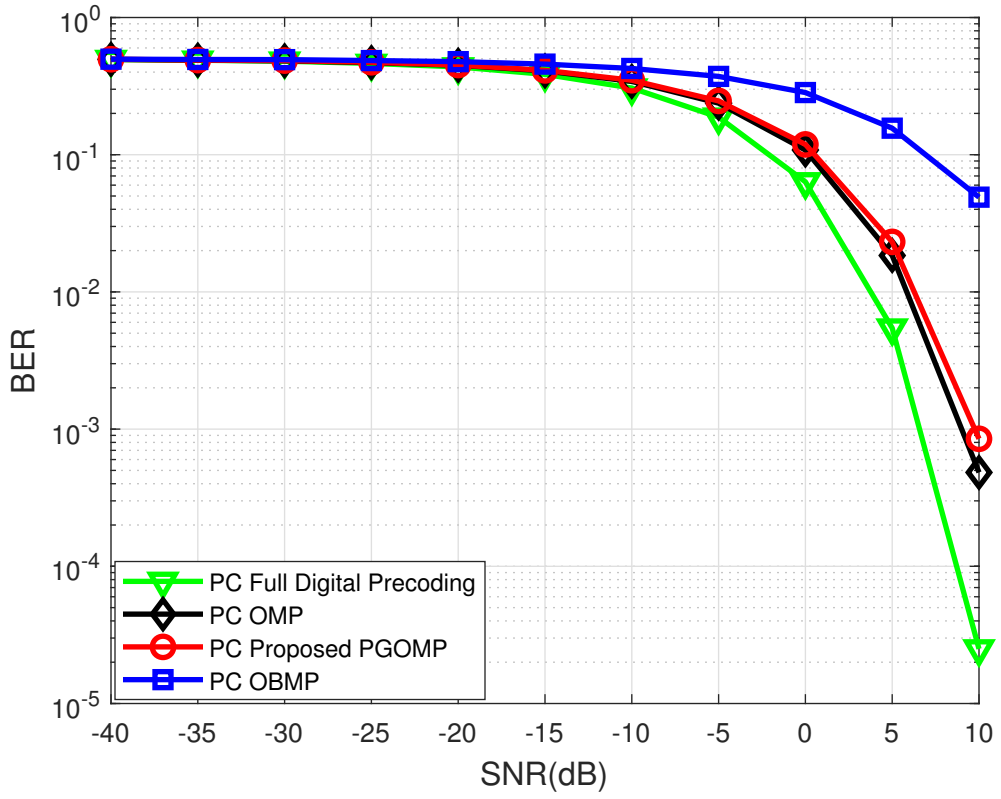

Figure 3.17: BER performance comparison vs SNR of different precoding algorithms with $N_{\mathrm{t}} = (16 \times 16) \times N_{\mathrm{s}}$, partially connected structure, $N_{\mathrm{r}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$.

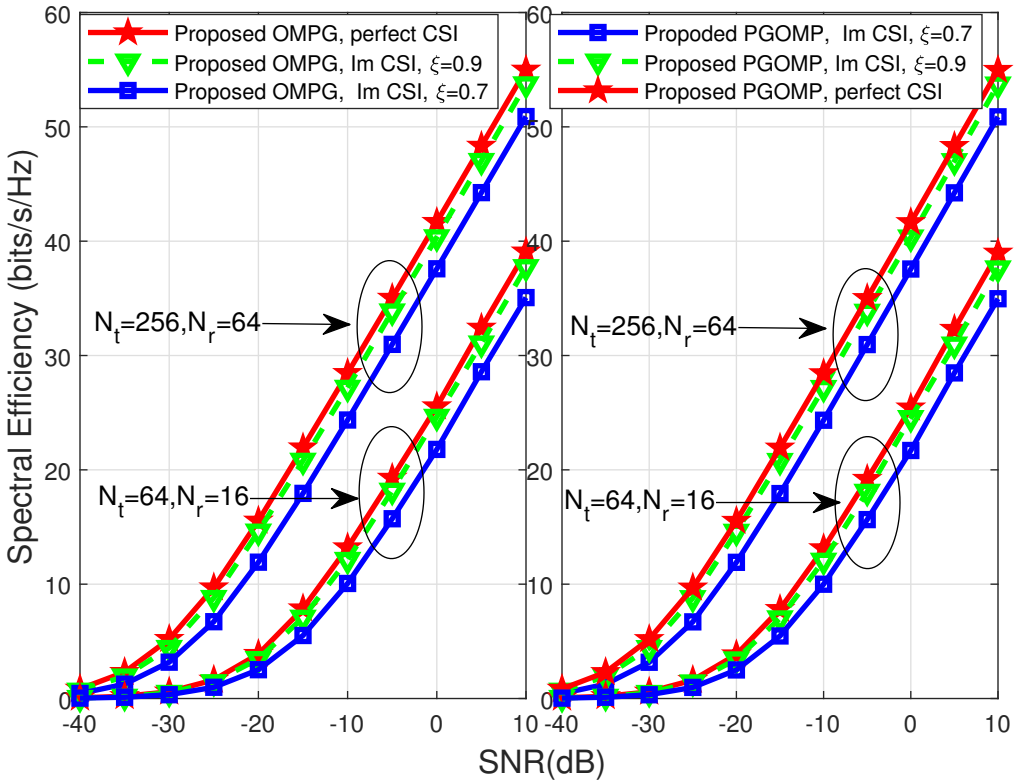

Figure 3.18: Impact of imperfect CSI on spectral efficiency performance, fully connected structure, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ and $N_{\mathrm{s}} = 4$.

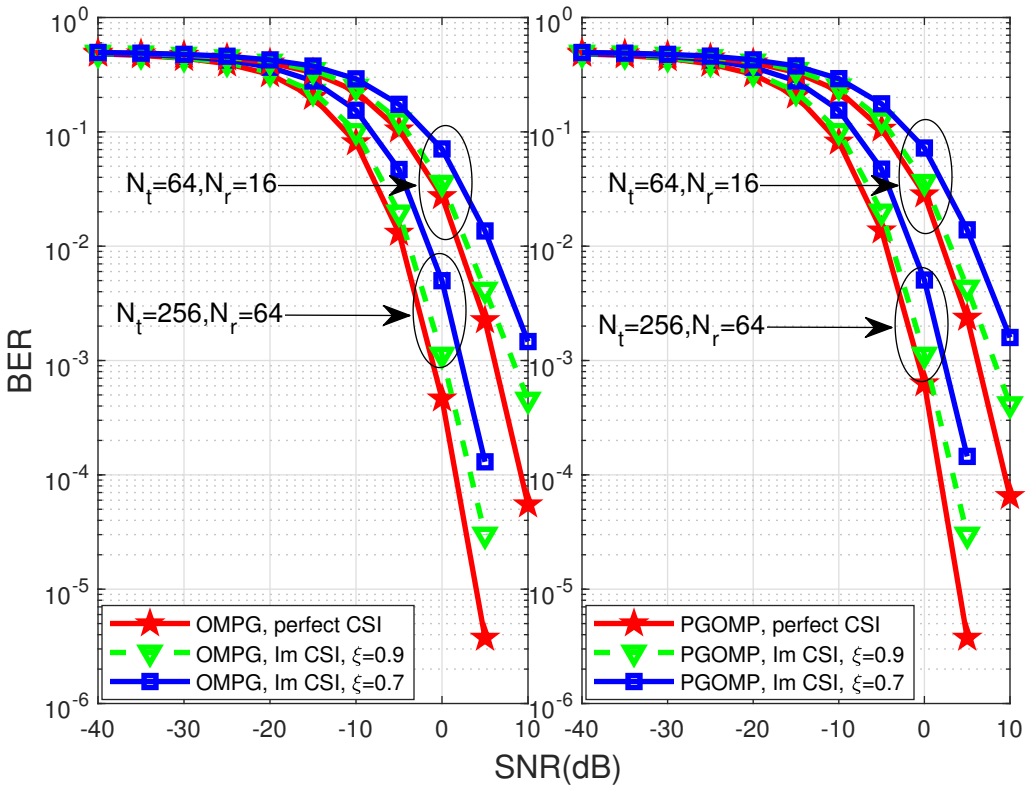

Figure 3.19: Impact of imperfect CSI on BER performance, fully connected structure, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 6$ and $N_{\mathrm{s}} = 4$.

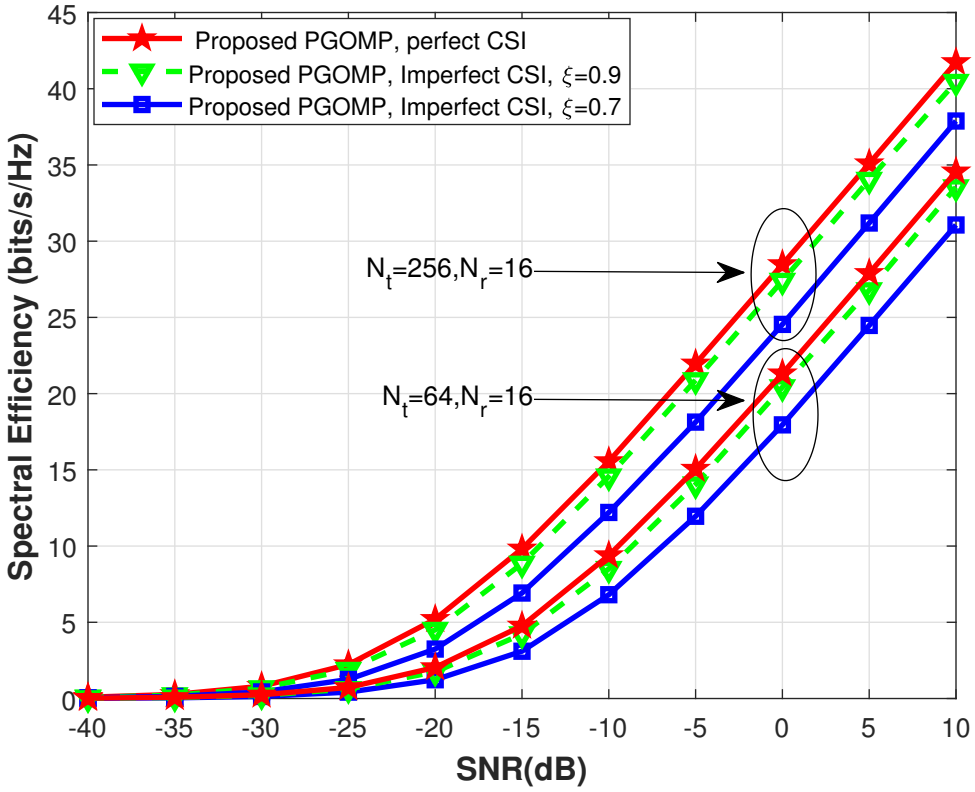

Figure 3.20: Impact of imperfect CSI on spectral efficiency performance, partially connected structure, $N_{\mathrm{r}} = 16$, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 4$ and $N_{\mathrm{s}} = 4$.

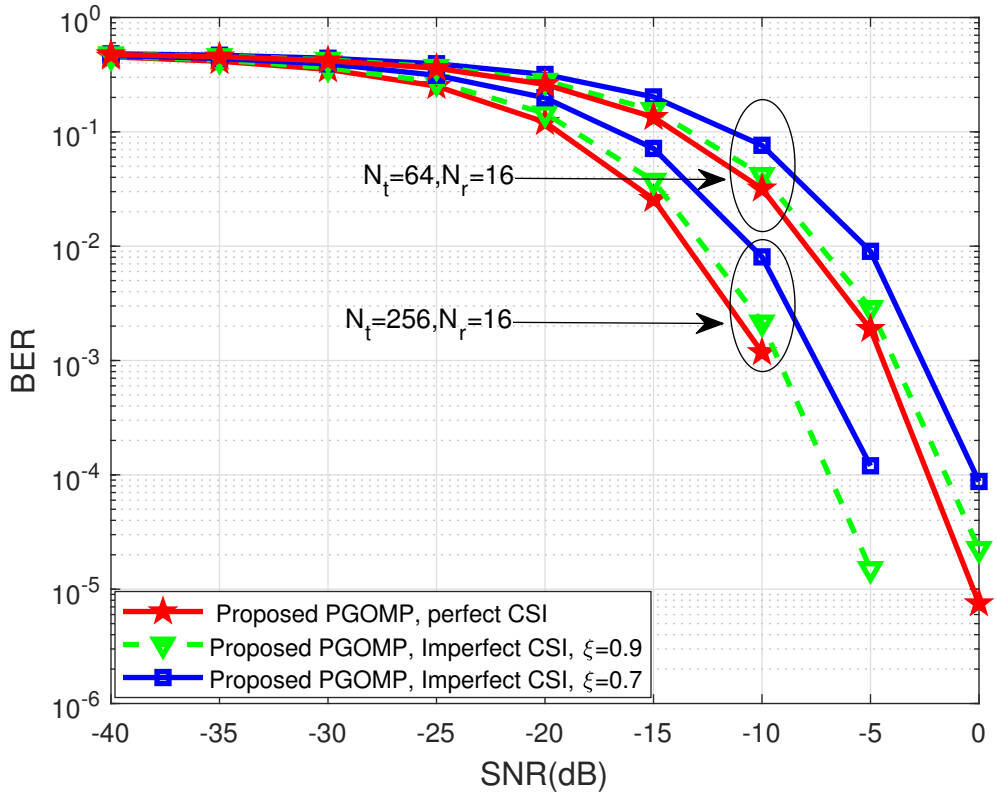

Figure 3.21: Impact of imperfect CSI on BER performance, partially connected structure, $N_{\mathrm{r}} = 16$, $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{RF}}^{\mathrm{r}} = 4$ and $N_{\mathrm{s}} = 4$.

Chapter 4

Interference Cancellation Aided Hybrid Beamforming for mmWave Multi-User Massive MIMO Systems

4.1 Introduction

In Chapter 3, we consider hybrid beamforming for point-to-point mmWave massive MIMO communication systems. As mmWave communication systems are expected to be deployed in small cell networks using massive MIMO antennas to provide gigabit data rates to users with high densities, hybrid beamforming is also studied for multi-user massive MIMO communication systems, where inter-user interference must be considered. In this chapter, we consider hybrid beamforming for mmWave MU-MIMO communications.

4.1.1 Prior Work and Motivations

In [60], the authors matched the RF precoder to the phase of the conjugate transpose of the composite downlink channel and designed the digital precoding by a zero-forcing (ZF) algorithm for the effective channel, where only single antenna users were considered. In [61], the authors proposed a two-stage hybrid beamforming algorithm. At the analog beamforming stage (or first stage), RF beamforming can be obtained by an exhaustive search algorithm over the candidate set constructed by beamsteering

codebooks with a similar format to array response vectors. At the digital precoding stage (or second stage), the conventional ZF precoding is applied to the equivalent baseband channel obtained from the first stage, and no digital combining is required. The authors in [59, 62, 100] also proposed a similar two-stage hybrid precoding algorithm. In [59], under the assumption that the number of RF chains at the BS is equal to the total number of RF chains of all MSs, the best RF combiners for each MS were obtained by an exhaustive search over DFT codebooks, and then found the RF precoder by equal gain transmission (EGT) to harvest the large array gain at the analog stage. At the digital stage, a baseband block diagonalization (BD) precoder [101] was designed to deal with inter-user interference and intra-user interference. However, the BD precoder suffers from performance loss due to the overlap of null space [101]. In [100], the authors proposed an OMP based algorithm to design hybrid MMSE precoder and combiners by minimizing the sum-MSE of the data streams intended for the MSs. In [62], the authors proposed to design the RF combiner by matching the left singular vectors of the downlink user channel and design the RF precoding by matching the right singular vectors of the equivalent channel at the analog stage. The BD scheme was then applied to the baseband equivalent channel to cancel inter-user interference at the digital stage. However, all of the aforementioned works are based on linear algorithms to cancel inter-user and intra-user interference, which is known to be suboptimal for multi-user MIMO downlink transmissions.

When considering multi-user and multi-stream downlink transmission, the optimal capacity can be achieved by nonlinear algorithms, i.e., dirty-paper coding (DPC) [102], Tomlinson-Harashima precoder (THP) [103], [104] and SIC [105], [106]. A hybrid design implementing the THP was proposed in [107]. However, it suffers from high complexity due to QR factorization of RF beamforming matrices and block-diagonal geometric mean decomposition (BD-GMD) of the effective channel matrix. SIC was proposed for multi-user signal detection in code division multiple access (CDMA) systems [105], [106], and was used to cancel multi-user interference in MU-MIMO systems [108], [109]. Regarding the existing literature about SIC-aided hybrid beamforming for mmWave massive MIMO systems, the authors in [64] proposed a hybrid precoding algorithm for fixed partially connected structures based on SIC. The paper uses the idea of SIC to decompose the optimization problem of total spectral efficiency into the sum of the spectral efficiency of the subarrays. However, the work only considers the single-user scenario. Inter-user interference and intra-user interference cancellation (IC) are not included. Inspired by [64], the authors

in [110] proposed MU-SIC joint hybrid precoding and combining design for partially connected architecture, but only used a linear BD algorithm to cancel inter-user and intra-user interference. An important contribution in this context is [111], in which the authors proposed an SIC-based uplink multiuser detection. However, the BS is only equipped with analog beamformers, which results in a limited performance compared to the BS with hybrid beamformers. Furthermore, each MS has a single antenna, which can only support one data stream transmission. More recently, a distributed digital and hybrid beamforming scheme with MMSE-SIC receivers was proposed for MIMO interference channels [112], which consists of multiple transmit-receive pairs with MIMO links. The authors developed a weighted MMSE (wMMSE) based SIC receiver to cancel intra-user interference, and linear spatial filters were used to cancel interference between different transmitter-receiver pairs.

From the discussion mentioned above, most of the existing works are focused on linear algorithms to cancel inter-user and intra-user interference. As non-linear IC algorithms can achieve better performance than linear ones, one direct motivation is to consider how to design non-linear IC-aided hybrid beamforming algorithms for downlink mmWave multi-user and multi-stream transmission.

In this chapter, we investigate the SIC-aided hybrid precoding and combining design for downlink mmWave multi-user massive MIMO systems, aiming to maximize the overall spectral efficiency. The rest of this chapter is organized as follows. In Section 4.2, we introduce the downlink mmWave multi-user massive MIMO hybrid beamforming system model and the formulated optimization problem. In Section 4.3, the proposed SIC-aided hybrid beamforming algorithms are presented. The computational complexity of the proposed algorithms is analyzed in Section 4.4. The simulation results used to evaluate the spectral efficiency performance of the proposed algorithms are provided in Section 4.5, followed by summary in Section 4.6. In this chapter, k refers to the kth user.

4.2 System Model and Problem Formulation

In this section, we introduce the hybrid beamforming structure of multi-user massive MIMO systems and mmWave channel model.

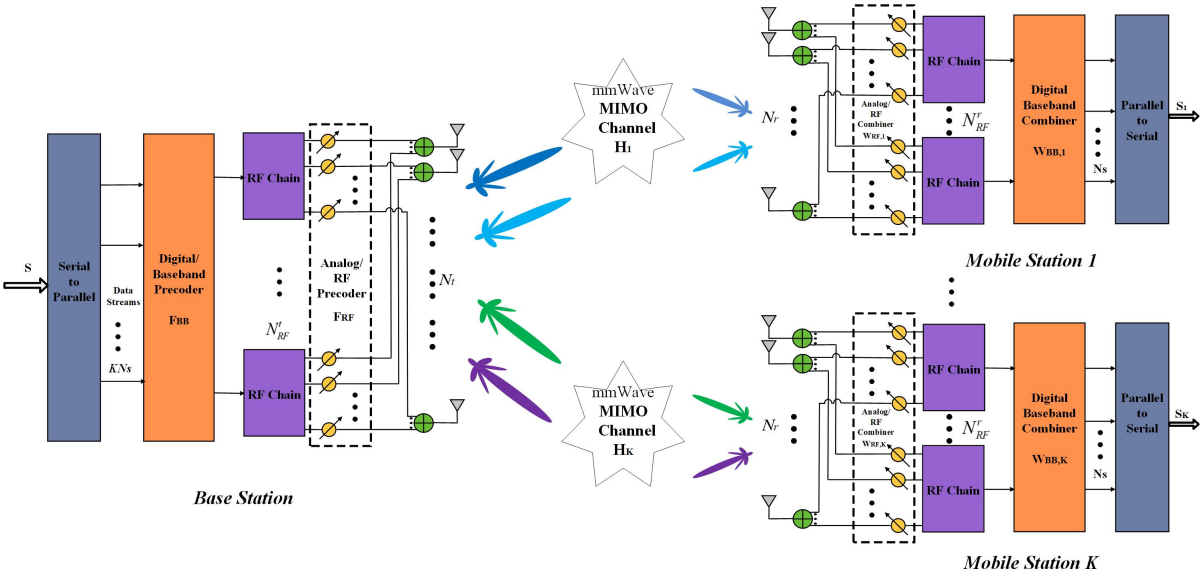

Figure 4.1: System diagram of a multi-user massive MIMO system with hybrid beamforming structure.

System Model

We consider the downlink multi-user communication of a massive MIMO system as shown in Figure 4.1 in which a BS with N_t massive antennas and N_{RF}^t limited RF chains is simultaneously communicating with K MSs. Each MS is equipped with N_r antennas and N_{RF}^r RF chains. Hence, each MS can support up to N_{RF}^r data streams. Without loss of generality, we assume each MS support N_s ($N_s \leq N_{RF}^r$) data streams, which means total KN_s ($KN_s \leq N_{RF}^t$) data streams are transmitted by the BS. To guarantee the effectiveness of the hybrid processing structure, the number of RF chains is constrained by $KN_s \leq N_{RF}^t \ll N_t$ for the BS and $N_s \leq N_{RF}^r \ll N_r$ for each MS. Note that to support N_s data streams transmission for each user, the least number of RF chains at each MS and the BS are $N_{RF}^r = N_s$ and $N_{RF}^t = KN_s = KN_{RF}^r$, respectively. In consideration of hardware cost and power consumption of RF chains, we assume the least number of RF chains in this chapter.

At the BS, this hybrid architecture enables the BS to apply an $N_{RF}^t \times KN_s$ digital baseband precoder $\mathbf{F}_{BB} = [\mathbf{F}_{BB,1}, \mathbf{F}_{BB,2}, \cdots, \mathbf{F}_{BB,K}]$, where $\mathbf{F}_{BB,i} \in \mathbb{C}^{N_{RF}^t \times N_s}$, followed by an $N_t \times N_{RF}^t$ fully-connected RF precoder $\mathbf{F}_{RF}$, which is implemented through an array of analog phase shifters. Since the phase shifters in the analog RF precoder (colored yellow) in Figure 4.1 can only change the phase of the transmitted signal, each entry of $\mathbf{F}_{RF}$ is of constant modulus. We normalize its entries to satisfy

$|\mathbf{F}_{RF}(i,j)| = \frac{1}{\sqrt{N_t}}, 1 \le i \le N_t, 1 \le j \le N_{RF}^t$. Furthermore, to meet the total transmit power constraints, $\mathbf{F}_{BB}$ is normalized to satisfy $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = KN_s$. The transmitted signal at the BS can be modeled as

$$\mathbf{X} = \mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{S} = \sum_{k=1}^{K} \mathbf{F}_{RF}\mathbf{F}_{BB,k}\mathbf{s}_k, \tag{4.1}$$

where $\mathbf{S} = \left[\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_K^T\right]^T \in \mathbb{C}^{KN_s \times 1}$ is the total transmitted symbols of all K users, and $\mathbf{s}_k \in \mathbb{C}^{N_s \times 1}$ denotes the transmitted symbol of user k. Assuming a narrowband block fading channel model, the received signal of the k-th MS is given by

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{X} + \mathbf{n}_k = \mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{S} + \mathbf{n}_k, \tag{4.2}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ represents the flat fading channel between MS k and the BS, $\mathbf{n}_k \in {}^{N_r \times 1}$ is the additive complex Gaussian noise at the k-th MS and each entry of $\mathbf{n}_k$ follows the independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and variance σ^2.

At the k-th MS, firstly the received signals are processed by a fully-connected RF combiner $\mathbf{W}_{RF,k} \in \mathbb{C}^{N_r \times N_{RF}^r}$ which is implemented by analog phase shifters similar to the BS. We also normalize its entries to satisfy $|\mathbf{W}_{RF,k}(i,j)| = \frac{1}{\sqrt{N_r}}, 1 \le i \le N_r, 1 \le j \le N_{RF}^r$. Then followed by a baseband combiner $\mathbf{W}_{BB,k}$ to detect its symbol. After combining, the detected signal at the k-th MS is given by

$$\begin{aligned}
\tilde{\mathbf{y}}_k = {} & \mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,i)s_{k,i} \\
& + \mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H\mathbf{H}_k \sum_{j=1,j\neq i}^{N_s} \mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,j)s_{k,j} \\
& + \mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H\mathbf{H}_k \sum_{l=1,l\neq k}^{K}\sum_{j=1}^{N_s} \mathbf{F}_{RF}\mathbf{F}_{BB,l}(:,j)s_{l,j} \\
& + \mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H\mathbf{n}_k.
\end{aligned} \tag{4.3}$$

The four successive terms of Equation (4.3) on different lines represent the i-th desired data stream for user k, the intra-user interference, the inter-user interference from other $K-1$ MSs and the effective noise, respectively. When the Gaussian symbol are utilized by the BS, the spectral efficiency achieved by the i-th data stream for the

k-th MS will be

$$R_{k,i} = \log_2 \left(|1 + SINR_{k,i}| \right). \tag{4.4}$$

The sum spectral efficiency is given by

$$R = \sum_{k=1}^{K} \sum_{i=1}^{N_s} R_{k,i} = \sum_{k=1}^{K} \sum_{i=1}^{N_s} \log_2 \left(|1 + SINR_{k,i}| \right), \tag{4.5}$$

where $SINR_{k,i} = \frac{P_s}{P_I + P_N}$ denotes SINR of the i-th stream of the k-th user. P_s, P_I and P_N are the desired signal power, interference power and noise power, which are given by Equations (4.6), (4.7) and (4.8), respectively. There, $P_{k,i}, k = 1, 2 \cdots K, i = 1, 2 \cdots N_s$ denotes the transmit power of the i-th stream of the k-th user.

$$P_s = P_{k,i} \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}(:,i) \mathbf{F}_{BB,k}^H(:,i) \mathbf{F}_{RF}^H \mathbf{H}_k^H \mathbf{W}_{RF,k} \mathbf{W}_{BB,k}(i,:) \tag{4.6}$$

$$\begin{aligned} P_I = {} & P_{k,j} \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{H}_k \sum_{j=1, j \neq i}^{N_s} \mathbf{F}_{RF} \mathbf{F}_{BB,k}(:,j) \mathbf{F}_{BB,k}^H(:,j) \mathbf{F}_{RF}^H \mathbf{H}_k^H \mathbf{W}_{RF,k} \mathbf{W}_{BB,k}(i,:) \\ & + P_{l,j} \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{H}_k \sum_{l=1, l \neq k}^{K} \sum_{j=1}^{N_s} \mathbf{F}_{RF} \mathbf{F}_{BB,l}(:,j) \mathbf{F}_{BB,l}^H(:,j) \mathbf{F}_{RF}^H \mathbf{H}_k^H \mathbf{W}_{RF,k} \mathbf{W}_{BB,k}(i,:) \end{aligned} \tag{4.7}$$

$$P_N = \sigma^2 \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{W}_{RF,k} \mathbf{W}_{BB,k}(i,:) \tag{4.8}$$

4.2.1 Channel Model

According to the mmWave channel model described in Chapter 2, the flat fading channel of the k-th MS, $\mathbf{H}_k$ can be written as

$$\mathbf{H}_k = \sqrt{\frac{N_t N_r}{N_C N_P}} \sum_{i=1}^{N_C} \sum_{l=1}^{N_P} \alpha_{il}^k \mathbf{a}_{MS}(\theta_{il}^k) \mathbf{a}_{BS}(\varphi_{il}^k)^H, \tag{4.9}$$

where $\alpha_{il}^k, i = 1, 2 \cdots N_c, l = 1, 2 \cdots N_P$ is the complex gain of the l-th ray in the i-th cluster of the k-th MS, which follows $\mathcal{CN}(0,1)$, while $\mathbf{a}_{\text{MS}}(\theta_{il}^k)$ and $\mathbf{a}_{\text{BS}}(\varphi_{il}^k)$ are the MS and BS array response vectors, respectively, where θ_{il}^k and φ_{il}^k denote the AoA and AoD of the l-th ray in the i-th cluster of the k-th MS, respectively. It is assumed that perfect CSI is available at both BS and MSs. In practical systems, CSI can be obtained by beamforming based channel estimation methods [113,114] or compressive sensing based channel estimation methods [115,116] and references therein.

4.2.2 Problem Formulation

The main purpose of this chapter is to maximize the sum spectral efficiency of all users under total transmit power and constant modulus constraints on each entry of the analog beamformers, assuming perfect knowledge of $\mathbf{H}_k$. That is, we aim to find the optimal hybrid precoder at the BS and optimal hybrid combiners for each user by solving the following problems:

$$\underset{\mathbf{F}_{RF},\mathbf{F}_{BB},\left\{\mathbf{W}_{RF,k},\mathbf{W}_{BB,k}\right\}_{k=1}^{K}}{\text{maximize}} \quad R \tag{4.10a}$$

$$\text{s.t.} \quad Tr\left(\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{F}_{BB}^{H}\mathbf{F}_{RF}^{H}\right) \leq P \tag{4.10b}$$

$$\left|\mathbf{F}_{RF}(i,j)\right|^2 = 1, \forall i,j \tag{4.10c}$$

$$\left|\mathbf{W}_{RF,k}(i,j)\right|^2 = 1, \forall i,j,k. \tag{4.10d}$$

To solve the above optimization problem as described in Equation (4.10a), a joint optimization over $\{\mathbf{F}_{RF}, \mathbf{F}_{BB}, \mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}\}$ is required. However, in general, the optimization problem described in Equation (4.10a) is a non-convex problem due to the presence of the variables $\{\mathbf{F}_{RF}, \mathbf{F}_{BB}\}$ and $\{\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}\}$ in the interference plus noise term $(P_I + P_N)$ and the product between the variables. In addition, as stated in [58], due to the non-convex constraints on $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF,k}$, finding the global optima for Equation (4.10a) is often found to be intractable. To simplify transceiver design, most of the aforementioned joint transmitter-receiver optimization problems in point-to-point mmWave massive MIMO communication systems are decoupled into the hybrid precoders $\{\mathbf{F}_{RF}, \mathbf{F}_{BB}\}$ design and the hybrid combiners $\{\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}\}$ design. For the hybrid precoders design, in lieu of maximizing spectral efficiency, the hybrid precoders $\{\mathbf{F}_{RF}, \mathbf{F}_{BB}\}$ are jointly designed to maximize the mutual information achieved by Gaussian signaling over the mmWave channel. Once the hybrid precoders are fixed, the hybrid combiners $\{\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}\}$ are jointly designed according to the same algorithm as the hybrid precoders. However, as mentioned before, most of these works cannot be directly extended to multi-user scenarios due to inter-user interference. Furthermore, the objective function in such a case is suboptimal compared to maximizing the sum spectral efficiency formulated in Equation (4.10a). For multi-user scenarios, the classic method is to divide the joint optimization problem into the RF stage and baseband stage optimization problems. At the RF stage, RF precoding and combining are alternately optimized to harvest array gain regardless of inter-user interference, and baseband precoding and combining are alternately

optimized at the baseband stage to deal with inter-user and intra-user interference. Recent research has shown that the separate RF and baseband design (also called as two-stage design) can achieve satisfactory performance without the need for myriad iterative procedures [59, 61, 62, 100]. Therefore, we use the two-stage design in this chapter.

4.3 SIC-Aided Hybrid Beamforming

In this section, we investigate SIC-aided hybrid beamforming design to cancel inter-user interference and intra-user interference under the proposed framework. From Equation (4.3) to enable inter-user interference-free and intra-user interference-free transmission, we need to remove inter-user and intra-user interference in turn. To the authors' best knowledge, most of the existing works use linear algorithms (ZF, MMSE or BD) to cancel inter-user interference and intra-user interference [59–62,100, 117–119]. However, linear algorithms suffer from performance loss due to the overlap of the null space of different users [101]. According to whether SIC is performed to cancel inter-user interference, intra-user interference, or both, we propose three SIC-aided hybrid beamforming designs which can be applied to three different scenarios. For the first design, we use the linear algorithm to cancel inter-user interference while SIC is adopted to cancel intra-user interference, which is suited to the scenario where intra-user interference is dominant. For the second design, we utilize SIC to cancel inter-user interference while the linear algorithm is used to cancel intra-user interference, which is suited to the scenario where inter-user interference is dominant. For the last design, SIC is utilized to suppress both inter-user and intra-user interference, and thus applicable to the scenario where both inter-user and intra-user interference are dominant. For simplicity, we choose the ZF algorithm for intra-user interference cancellation and the BD algorithm for inter-user interference cancellation to demonstrate that the SIC algorithm outperforms its linear counterpart. It should be noted the proposed IC framework in this chapter is general for any linear IC algorithms. The number of users and the number of data streams per user can be obtained by system parameters. Therefore, the relative dominance of the inter-user and intra-user interference can be obtained. For example, during the system scheduling step, far apart users are typically paired together as groups, and these users between different groups will be spatially separated and thus have minimal inter-user interference. In each user group, the relative dominance of the inter-user and intra-user interference

can be determined according to the number of users and the number of data streams per user.

For all the three designs, we focus on the baseband precoder and combiner. The RF precoder and combiner can be obtained through the codebook exhaustive search [61], array gain harvesting algorithms [59], or phase extraction [62]. Motivated by maximizing the capacity of the baseband channel and acceptable complexity of multi-user hybrid RF beamforming algorithm, we use the phase extraction design for RF precoder and combiner [62]. Note that the proposed algorithms in this chapter are general for any RF precoder and combiner.

4.3.1 RF Beamforming

When the double of the least number of RF chains are available at transceivers, i.e., $N_{RF}^t = 2KN_s, N_{RF}^r = 2N_s$, two RF chains with the constant modulus constraints can be combined to act as one RF chain without the constant modulus constraint, the RF combiner of the k-th MS, $\mathbf{W}_{RF,k} = \mathbf{U}_k^{\left(N_{RF}^r\right)}$, where $\mathbf{U}_k^{\left(N_{RF}^r\right)}$ consists of the first N_{RF}^r left singular vectors of $\mathbf{H}_k$. The RF precoder, $\mathbf{F}_{RF} = \mathbf{V}_{int}^{\left(N_{RF}^t\right)}$, where $\mathbf{V}_{int}^{\left(N_{RF}^t\right)}$ consists of the first N_{RF}^t right singular vectors of $\mathbf{H}_{int}$, where $\mathbf{H}_{int}$ is given by

$$\mathbf{H}_{int} = \begin{bmatrix} \mathbf{W}_{RF,1}^H \mathbf{H}_1 \\ \vdots \\ \mathbf{W}_{RF,K}^H \mathbf{H}_K \end{bmatrix}_{KN_{RF}^r \times N_t}. \tag{4.11}$$

When the double of the least number of RF chains are not available at transceivers, i.e., $KN_s \leq N_{RF}^t < 2KN_s, N_s \leq N_{RF}^r < 2N_s$, $\mathbf{W}_{RF,k}$ and $\mathbf{F}_{RF}$ can be obtained by extracting the phase of $\mathbf{U}_k^{\left(N_{RF}^r\right)}$ and $\mathbf{V}_{int}^{\left(N_{RF}^t\right)}$, respectively, which are given by

$$\mathbf{W}_{RF,k} = \frac{1}{\sqrt{N_r}} e^{j\angle \mathbf{U}_k^{\left(N_{RF}^r\right)}}. \tag{4.12}$$

$$\mathbf{F}_{RF} = \frac{1}{\sqrt{N_t}} e^{j\angle \mathbf{V}_{\text{int}}^{\left(N_{RF}^t\right)}}. \tag{4.13}$$

The baseband equivalent channel of the k-th MS $\tilde{\mathbf{H}}_k$ with dimension $N_{RF}^r \times N_{RF}^t$ is

given by

$$\tilde{\mathbf{H}}_k = \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF}, k = 1, 2, \cdots, K. \tag{4.14}$$

4.3.2 Inter-user IC with Linear Algorithm, Intra-user IC with SIC

Baseband Precoder In this case, we use the linear algorithm to remove inter-user interference and adopt SIC to cancel intra-user interference. We define $\bar{\mathbf{H}}_k$ as

$$\bar{\mathbf{H}}_k = \left[\tilde{\mathbf{H}}_1^T, \cdots, \tilde{\mathbf{H}}_{k-1}^T, \tilde{\mathbf{H}}_{k+1}^T, \cdots, \tilde{\mathbf{H}}_K^T\right]^T. \tag{4.15}$$

After SVD of $\bar{\mathbf{H}}_k$, $\bar{\mathbf{V}}_k^{\left(N_{RF}^r\right)}$ which consists of the last N_{RF}^r right singular vectors of $\bar{\mathbf{H}}_k$ can be acquired. From the definition of SVD, the columns of $\bar{\mathbf{V}}_k^{\left(N_{RF}^r\right)}$ are exactly the orthogonal bases of the null space of $\bar{\mathbf{H}}_k$. To cancel the inter-user interference, we force the baseband precoder of the k-th MS $\mathbf{F}_{BB,k}$ to lie in the null space of $\bar{\mathbf{H}}_k$, given by

$$\mathbf{F}_{BB,k} = \bar{\mathbf{V}}_k^{\left(N_{RF}^r\right)}. \tag{4.16}$$

Note that the baseband precoder is different from the counterpart of BD [59]. Furthermore, it can be seen from Equation (4.15) and (4.16), the null space of different MSs are overlapped, and therefore, BD suffers from performance loss [101]. By using the baseband precoder as given by Equation (4.16), the k-th MS does not see any interference from the other $K-1$ MSs. The overall baseband precoder with the assumption of $N_{RF}^t = KN_{RF}^r$ is given by

$$\begin{aligned} \mathbf{F}_{BB} &= [\mathbf{F}_{BB,1}, \mathbf{F}_{BB,2}, \cdots, \mathbf{F}_{BB,K}] \\ &= \left[\bar{\mathbf{V}}_1^{\left(N_{RF}^r\right)}, \bar{\mathbf{V}}_2^{\left(N_{RF}^r\right)}, \cdots, \bar{\mathbf{V}}_K^{\left(N_{RF}^r\right)}\right]. \end{aligned} \tag{4.17}$$

The baseband precoder of the k-th MS $\mathbf{F}_{BB,k}$ satisfies:

$$\begin{aligned} \tilde{\mathbf{H}}_k \mathbf{F}_{BB,l} &= \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,l} \\ &= \begin{cases} \mathbf{0}, & l \neq k \\ \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \bar{\mathbf{V}}_k^{\left(N_{RF}^r\right)}, & l = k \end{cases} \end{aligned} \tag{4.18}$$

In this case, the received signal of the k-th MS after RF combining is given by

$$\begin{aligned}\mathbf{y}_{RF,k} &= \mathbf{W}_{RF,k}^{H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,i)s_{k,i} \\ &\quad + \sum_{j=1,j\neq i}^{N_s}\mathbf{W}_{RF,k}^{H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,j)s_{k,j} \\ &\quad + \mathbf{W}_{RF,k}^{H}\mathbf{n}_k.\end{aligned} \tag{4.19}$$

The three successive terms of Equation (4.19) on different lines represent the i-th desired data stream for user k, the intra-user interference and the effective noise, respectively. It can be seen from Equation (4.19), the inter-user interference has been cancelled due to Equation (4.18).

Baseband Combiner The SIC-aided baseband combiner is designed to cancel intra-user interference at the receiver side. When SIC is used, the order in which streams of the k-th MS $\mathbf{s}_k$ are detected becomes important to the overall performance of the system. For now, we first discuss the general detection procedure with respect to arbitrary ordering. For the k-th MS, define the equivalent channel as

$$\hat{\mathbf{H}}_k = \mathbf{W}_{RF,k}^{H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k} = \left[\hat{\mathbf{h}}_{k,1},\cdots,\hat{\mathbf{h}}_{k,N_s}\right]. \tag{4.20}$$

To perform SIC between multi-streams of the k-th MS, define $\mathbf{G}_k = \left(\hat{\mathbf{H}}_k^{H}\hat{\mathbf{H}}_k\right)^{-1}\hat{\mathbf{H}}_k^{H}$. The baseband combiner to detect the i-th data stream of the k-th MS is $\mathbf{W}_{BB,k}^{H}(i,:) = \mathbf{G}_k(i,:)$. The detected i-th data stream of the k-th MS $\hat{s}_{k,i}$ is given by

$$\begin{aligned}\hat{s}_{k,i} &= \mathbf{W}_{BB,k}^{H}(i,:)\mathbf{y}_{RF,k} \\ &= \mathbf{W}_{BB,k}^{H}(i,:)\mathbf{W}_{RF,k}^{H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,i)s_{k,i} \\ &\quad + \mathbf{W}_{BB,k}^{H}(i,:)\mathbf{W}_{RF,k}^{H}\mathbf{n}_k.\end{aligned} \tag{4.21}$$

The two successive terms of the second equality of Equation (4.21) on different lines represent the i-th desired data stream for user k and the effective noise, respectively. It can be seen from Equation (4.21), the intra-user interference has been cancelled due to $\mathbf{W}_{BB,k}^{H}(i,:) = \mathbf{G}_k(i,:)$. Therefore, we can detect the i-th data stream of the k-th MS $\hat{s}_{k,i}$, and then $\hat{s}_{k,i}$ is demodulated and re-modulated to cancel the noise accumulation. After that the re-modulated symbol $\tilde{s}_{k,i}$ will be subtracted from the

received signal, which is given by

$$\begin{aligned}\bar{\mathbf{y}}_{k,i} &= \mathbf{y}_{RF,k} - \hat{\mathbf{h}}_{k,i}\tilde{s}_{k,i} \\ &= \sum_{j=1,j\neq i}^{N_s} \tilde{\mathbf{H}}_k \mathbf{F}_{BB,k}(:,j)s_{k,j} + \mathbf{W}_{RF,k}^H \mathbf{n}_k.\end{aligned} \tag{4.22}$$

Next, we will detect the n-th ($n \neq i$) data stream of the k-th MS $s_{k,n}$. Let $\bar{\mathbf{H}}_{k,i}$ denote zeroing column i of $\hat{\mathbf{H}}_k$, which is given by

$$\bar{\mathbf{H}}_{k,i} = \left[\hat{\mathbf{h}}_{k,1}, \cdots, \hat{\mathbf{h}}_{k,i-1}, \mathbf{0}, \hat{\mathbf{h}}_{k,i+1}, \cdots, \hat{\mathbf{h}}_{k,N_s}\right]. \tag{4.23}$$

Define $\bar{\mathbf{G}}_{k,i} = \left(\bar{\mathbf{H}}_{k,i}^H \bar{\mathbf{H}}_{k,i}\right)^{-1} \bar{\mathbf{H}}_{k,i}^H$. The baseband combiner to detect the n-th data stream of the k-th MS $s_{k,n}$ is $\mathbf{W}_{BB,k}^H(n,:) = \bar{\mathbf{G}}_{k,i}(n,:)$, and the detected n-th data stream of the k-th MS $\hat{s}_{k,n}$ is given by

$$\begin{aligned}\hat{s}_{k,n} &= \mathbf{W}_{BB,k}^H(n,:)\bar{\mathbf{y}}_{k,i} \\ &\overset{(a)}{=} \mathbf{W}_{BB,k}^H(n,:)\mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF}\mathbf{F}_{BB,k}(:,n)s_{k,n} \\ &\quad + \mathbf{W}_{BB,k}^H(n,:)\mathbf{W}_{RF,k}^H \mathbf{n}_k\ .\end{aligned} \tag{4.24}$$

The simplification step (a) of Equation (4.24) holds due to $\mathbf{W}_{BB,k}^H(n,:) = \bar{\mathbf{G}}_{k,i}(n,:)$. Then the detected symbol $s_{k,i}$ and $s_{k,n}$ will be subtracted from the received signal, and the above steps are performed for the rest streams by operating in turn on the progression of the detected symbol removed.

In summary, we provide inter-user IC with the linear algorithm and intra-user IC with SIC in Algorithm 6. To distinguish, we refer to Algorithm 6 as the HBF SIC I Algorithm.

Detection Stream Order As mentioned before, when SIC is used, the order in which streams of the k-th MS's symbol $\mathbf{s}_k$ are detected becomes important to the overall performance of the system. We choose the best post-detection SINR of the i-th stream of the k-th MS at each stream detection, which is given by Equation (4.25). Note that P_I in Equation (4.26) is different from Equation (4.7) since the inter-user interference has been cancelled in Equation (4.19). The SINR can be obtained by user measurements.

$$SINR_{k,i} = \frac{P_{k,i}\mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF}\mathbf{F}_{BB,k}(i,:)(\mathbf{F}_{BB,k}(i,:))^H \mathbf{F}_{RF}^H \mathbf{H}_k^H \mathbf{W}_{RF,k}\mathbf{W}_{BB,k}(i,:)}{\sigma^2 \mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H \mathbf{W}_{RF,k}\mathbf{W}_{BB,k}(i,:) + P_I} \tag{4.25}$$

Algorithm 6 The HBF SIC I Algorithm

Input: $\mathbf{H}_k, 1 \leq k \leq K$
Analog Stage
1. **for** $k \leq K$ **do**
2. Calculate $\mathbf{W}_{RF,k}$ according to Equation (4.12)
3. Calculate $\mathbf{H}_{int}$ according to Equation (4.11)
4. **end for**
5. Calculate $\mathbf{F}_{RF}$ according to Equation (4.13)
6. Calculate $\tilde{\mathbf{H}}_k, k = 1, 2 \cdots, K$, according to Equation (4.14)
End Analog Stage
Digital Stage
7. **for** $k \leq K$ **do**
8. $\bar{\mathbf{H}}_k = \left[\tilde{\mathbf{H}}_1^T, \cdots, \tilde{\mathbf{H}}_{k-1}^T, \tilde{\mathbf{H}}_{k+1}^T, \cdots, \tilde{\mathbf{H}}_K^T\right]$
9. $\bar{\mathbf{H}}_k = \bar{\mathbf{U}}_k \bar{\mathbf{\Sigma}}_k \left[\bar{\mathbf{V}}_k^{((K-1)N_{RF}^r)}, \bar{\mathbf{V}}_k^{(N_{RF}^r)}\right]^H$
10. $\mathbf{F}_{BB,k} = \bar{\mathbf{V}}_k^{(N_{RF}^r)}$
11. **end for**
12. **Obtain** $\mathbf{F}_{BB} = [\mathbf{F}_{BB,1}, \cdots, \mathbf{F}_{BB,K}]$
13. **for** $k \leq K$ **do**
14. $\hat{\mathbf{H}}_k = \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB}$
15. **for** $i \leq N_s$ **do**
16. $\bar{\mathbf{H}}_{k,i} = \left[\tilde{\mathbf{h}}_{k,1}, \cdots, \tilde{\mathbf{h}}_{k,i-1}, \mathbf{0}, \tilde{\mathbf{h}}_{k,i+1}, \tilde{\mathbf{h}}_{k,N_s}\right]$
17. $\bar{\mathbf{G}}_{k,i} = \left(\bar{\mathbf{H}}_{k,i}^H \bar{\mathbf{H}}_{k,i}\right)^{-1} \bar{\mathbf{H}}_{k,i}^H$
18. $\mathbf{W}_{BB,k}(i,:) = \bar{\mathbf{G}}_{k,i}(i,:)$
19. **end for**
20. **end for**
End Digital Stage
Output $\mathbf{F}_{RF}, \mathbf{F}_{BB}, (\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}), k = 1 : K$

$$P_I = P_{k,j}\mathbf{W}_{BB,k}^H(i,:)\mathbf{W}_{RF,k}^H\mathbf{H}_k\mathbf{F}_{RF}\sum_{j=1,j\neq i}^{N_s}\mathbf{F}_{BB,k}(j,:)(\mathbf{F}_{BB,k}(j,:))^H\mathbf{F}_{RF}^H\mathbf{H}_k^H\mathbf{W}_{RF,k}\mathbf{W}_{BB,k}(i,:) \tag{4.26}$$

4.3.3 Inter-user IC with SIC, Intra-user IC with Linear Algorithm

Baseband Precoder In this case, linear successive optimization (SO) at the BS [101] and non-linear SIC at the MS are used to remove inter-user interference, and ZF is adopted to cancel intra-user interference. As mentioned before, BD suffers from performance loss due to the overlap of null space. To address this issue, we utilize SO implemented by baseband precoder and SIC implemented after RF combiner to remove inter-user interference. That is to say, if we want to detect the k-th MS, the previous $k-1$ MSs are suppressed by SO, and the rest $k+1$ to K-th MS are cancelled by SIC. To cancel the inter-user interference due to the previous $k-1$ MSs, the baseband precoder of the l-th MS $\mathbf{F}_{BB,l}^{SO}$ needs to satisfy $\tilde{\mathbf{H}}_k\mathbf{F}_{BB,l}^{SO} = \mathbf{0}, 1 \leq l < k$, and we define $\bar{\mathbf{H}}_k^{SIC}, (1 \leq k < K)$ different from BD, written as

$$\bar{\mathbf{H}}_k^{SIC} = \left[\tilde{\mathbf{H}}_{k+1}^T, \tilde{\mathbf{H}}_{k+2}^T, \cdots, \tilde{\mathbf{H}}_K^T\right]^T. \tag{4.27}$$

After SVD of $\bar{\mathbf{H}}_k^{SIC}$, $\hat{\mathbf{V}}_k^{\left(N_{RF}^t-(K-k)N_{RF}^r\right)}$ which consists of the last $N_{RF}^t-(K-k)\,N_{RF}^r$ right singular vectors of $\bar{\mathbf{H}}_k^{SIC}$ can be obtained. From the definition of SVD, the columns of $\hat{\mathbf{V}}_k^{\left(N_{RF}^t-(K-k)N_{RF}^r\right)}$ are exactly the orthogonal bases of the null space of $\bar{\mathbf{H}}_k^{SIC}$. Comparing Equation (4.27) with Equation (4.15), we can observe the null space of $\bar{\mathbf{H}}_k^{SIC}, 1 \leq k < K$ are not overlapped any more, and therefore, the SIC based inter-user cancellation algorithm outperforms its linear counterparts. For the K-th MS, there is no interfering user according to our design. After performing SVD of its baseband equivalent channel $\tilde{\mathbf{H}}_K$, $\hat{\mathbf{V}}_K^{\left(N_{RF}^r\right)}$ which consists of the first N_{RF}^r right singular vectors of $\tilde{\mathbf{H}}_K$ can be obtained. Therefore, the K-th MS can perform inter-user interference-free multi-stream transmission.

For the first $K-1$ MSs, to achieve inter-user interference-free transmission, we force the baseband precoder $\mathbf{F}_{BB,k}^{SO}$ of the k-th MS lies in the null space of $\bar{\mathbf{H}}_k^{SIC}$ by setting $\mathbf{F}_{BB,k}^{SO} = \hat{\mathbf{V}}_k^{(0)}\mathbf{G}_k, 1 \leq k < K$, where $\hat{\mathbf{V}}_k^{(0)} = \hat{\mathbf{V}}_k^{\left(N_{RF}^t-(K-k)N_{RF}^r\right)}$ denotes the null space of $\bar{\mathbf{H}}_k^{SIC}$. For the K-th MS, $\mathbf{F}_{BB,K}^{SO} = \hat{\mathbf{V}}_K^{(N_{RF}^r)}\mathbf{G}_K$, where $\mathbf{G}_k, 1 \leq k \leq K$ is

derived to maximize the spectral efficiency of the k-th MS, which is given by

$$\mathbf{G}_k = \mathbf{M}_k^{(N_s)} \mathbf{\Gamma}_k^{1/2}, \tag{4.28}$$

Proof: See Appendix A or [65].

The baseband precoder $\mathbf{F}_{BB,k}^{SO}$ is given by

$$\mathbf{F}_{BB,k}^{SO} = \begin{cases} \hat{\mathbf{V}}_k^{\left(N_{RF}^t - (K-k)N_{RF}^r\right)} \mathbf{G}_k & 1 \leq k < K \\ \hat{\mathbf{V}}_K^{\left(N_{RF}^r\right)} \mathbf{G}_K & k = K \end{cases} \tag{4.29}$$

Finally, the overall baseband precoder to enable inter-user interference-free transmission is given by

$$\mathbf{F}_{BB}^{SO} = \left[\mathbf{F}_{BB,1}^{SO}, \mathbf{F}_{BB,2}^{SO}, \cdots, \mathbf{F}_{BB,K}^{SO}\right]. \tag{4.30}$$

From the previous derivation, the baseband precoder of the k-th MS $\mathbf{F}_{BB,k}^{SIC}$ satisfies:

$$\begin{aligned} \tilde{\mathbf{H}}_k \mathbf{F}_{BB,l}^{SO} &= \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,l}^{SO} \\ &= \begin{cases} \mathbf{0} & 1 \leq l < k \\ \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,l}^{SO} \hat{\mathbf{V}}_k^{(0)} \mathbf{G}_k & k \leq l < K-1 \\ \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,K}^{SO} \hat{\mathbf{V}}_K^{\left(N_{RF}^r\right)} \mathbf{G}_K & l = K \end{cases} \end{aligned} \tag{4.31}$$

It can be seen from Equation (4.31), for the k-th MS, the previous $k-1$ MSs has been cancelled due to the linear SO based inter-user cancellation baseband precoding as given by Equation (4.29) and (4.30). After RF combining, the received signal of the k-th MS is given by

$$\begin{aligned} \mathbf{y}_{RF,k} &= \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB}^{SO} \mathbf{S} + \mathbf{W}_{RF,k}^H \mathbf{n}_k \\ &\overset{(b)}{=} \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO} \mathbf{s}_k \\ &\quad + \mathbf{W}_{RF,k}^H \mathbf{H}_k \sum_{l=k+1}^{K} \mathbf{F}_{RF} \mathbf{F}_{BB,l}^{SO} \mathbf{s}_l \\ &\quad + \mathbf{W}_{RF,k}^H \mathbf{n}_k. \end{aligned} \tag{4.32}$$

The simplification step (b) of Equation (4.32) holds due to Equation (4.31). So far, all the previous $k-1$ MSs have been cancelled. To achieve inter-user interference-free transmission, the $(k+1)$-th to K-th MS are required to be subtracted from the

received signal, which is given by

$$\begin{aligned}\bar{\mathbf{y}}_k &= \mathbf{y}_{RF,k} - \sum_{l=k+1}^{K} \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,l}^{SO} \mathbf{s}_l \\ &= \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO} \mathbf{s}_k + \mathbf{W}_{RF,k}^H \mathbf{n}_k.\end{aligned} \tag{4.33}$$

Note that, due to SIC implemented after RF combiner at the receiver side, the detected $k+1$ to K-th MS signals are removed from the received signal of the k-th MS.

Baseband Combiner Then we perform baseband combining with $\mathbf{W}_{BB,k}$ of the inter-user interference-free signal, and the ZF baseband combiner is given by

$$\mathbf{W}_{BB,k} = \left(\mathbf{H}_{Equ,k}^H \mathbf{H}_{Equ,k}\right)^{-1} \mathbf{H}_{Equ,k}^H, \tag{4.34}$$

where $\mathbf{H}_{Equ,k} = \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO}$ denotes the equivalent channel of the kth MS. The baseband combined signal is given by

$$\hat{\mathbf{s}}_k = \mathbf{W}_{BB,k}^H \bar{\mathbf{y}}_k \overset{(c)}{=} \mathbf{s}_k + \mathbf{W}_{BB,k}^H \mathbf{W}_{RF,k}^H \mathbf{n}_k. \tag{4.35}$$

The simplification step (c) of Equation (4.35) holds due to ZF baseband combiner given in Equation (4.34), i.e., $\mathbf{W}_{BB,k}^H \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO} = \mathbf{I}_{N_s}$, which results in inter-user interference-free and intra-user interference-free transmission of the k-th MS. Without considering the detection order of MSs, the MSs will be detected in the reverse order, i.e, the K-th MS will be detected first. According to Equation (4.35), $\hat{\mathbf{s}}_K$ is given by

$$\hat{\mathbf{s}}_K = \mathbf{s}_K + \mathbf{W}_{BB,K}^H \mathbf{W}_{RF,K}^H \mathbf{n}_K. \tag{4.36}$$

To detect the $(K-1)$-th MS, the detected K-th MS will be subtracted from the received RF combined signal of the $(K-1)$-th MS, which is given by

$$\begin{aligned}\bar{\mathbf{y}}_{K-1} &= \mathbf{y}_{RF,K-1} - \tilde{\mathbf{H}}_{K-1} \mathbf{F}_{BB,K}^{SO} \mathbf{s}_K \\ &= \mathbf{W}_{RF,K-1}^H \mathbf{H}_{K-1} \mathbf{F}_{RF} \mathbf{F}_{BB,K-1}^{SO} \mathbf{s}_{K-1} \\ &\quad + \mathbf{W}_{RF,K-1}^H \mathbf{n}_{K-1}.\end{aligned} \tag{4.37}$$

Algorithm 7 The HBF SIC II Algorithm

Input: $\mathbf{H}_k, 1 \leq k \leq K$
Analog Stage
1. **for** $k \leq K$ **do**
2. Calculate $\mathbf{W}_{RF,k}$ according to Equation (4.12)
3. Calculate $\mathbf{H}_{int}$ according to Equation (4.11)
4. **end for**
5. Calculate $\mathbf{F}_{RF}$ according to Equation (4.13)
6. Calculate $\tilde{\mathbf{H}}_k, k = 1, 2 \cdots, K$, according to Equation (4.14)
End Analog Stage
Digital Stage
7. **for** $k \leq K$ **do**
8. **if** $k < K$ **do**
9. $\bar{\mathbf{H}}_k^{SIC} = \left[\tilde{\mathbf{H}}_{k+1}^T, \tilde{\mathbf{H}}_{k+2}^T, \cdots, \tilde{\mathbf{H}}_K^T\right]^T$
10. $\bar{\mathbf{H}}_k^{SIC} = \hat{\mathbf{U}}_k \hat{\mathbf{\Sigma}}_k \hat{\mathbf{V}}_k^H$
11. $\mathbf{F}_{BB,k}^{SO} = \hat{\mathbf{V}}_k^{\left(N_{RF}^t - (K-k)N_{RF}^r\right)} \mathbf{G}_k$
12. $\mathbf{H}_{Equ,k} = \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO}$
13. $\mathbf{W}_{BB,k} = \left(\mathbf{H}_{Equ,k}^H \mathbf{H}_{Equ,k}\right)^{-1} \mathbf{H}_{Equ,k}^H$
14. **else**
15. $\tilde{\mathbf{H}}_K = \hat{\mathbf{U}}_K \hat{\mathbf{\Sigma}}_K \left[\hat{\mathbf{V}}_K^{(N_{RF}^r)}, \hat{\mathbf{V}}_K^{\left(N_{RF}^t - N_{RF}^r\right)}\right]^H$
16. $\mathbf{F}_{BB,K}^{SO} = \hat{\mathbf{V}}_K^{(N_{RF}^r)} \mathbf{G}_K$
17. $\mathbf{H}_{Equ,K} = \mathbf{W}_{RF,K}^H \mathbf{H}_K \mathbf{F}_{RF} \mathbf{F}_{BB,K}^{SO}$
18. $\mathbf{W}_{BB,K} = \left(\mathbf{H}_{Equ,K}^H \mathbf{H}_{Equ,K}\right)^{-1} \mathbf{H}_{Equ,K}^H$
19. **end if**
20. **end for**
21. **Obtain** $\mathbf{F}_{BB}^{SO} = \left[\mathbf{F}_{BB,1}^{SO}, \cdots, \mathbf{F}_{BB,K}^{SO}\right]$
End Digital Stage
Output $\mathbf{F}_{RF}, \mathbf{F}_{BB}^{SO}, \left(\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k}\right), k = 1 : K$

After ZF baseband combining with $\mathbf{W}_{BB,K-1}$, the obtained signal is given by

$$\begin{aligned} \hat{\mathbf{s}}_{K-1} &= \mathbf{W}_{BB,K-1}^H \bar{\mathbf{y}}_{K-1} \\ &= \mathbf{s}_{K-1} + \mathbf{W}_{BB,K-1}^H \mathbf{W}_{RF,K-1}^H \mathbf{n}_{K-1}. \end{aligned} \tag{4.38}$$

Obviously, the $(K-1)$ th MS can be detected from Equation (4.38), the rest MSs are performed by operating in turn on the progression of the detected MSs removed. Therefore, the number of interference users to the later-detect user is less than the previous ones. The order in which users are detected can be determined by their large scale fading, for example, near MSs with less path attenuation should be detected first. In summary, we provide inter-user IC with SIC and intra-user IC with ZF in Algorithm 7. To distinguish, we refer to Algorithm 7 as the HBF SIC II Algorithm.

4.3.4 Both Inter-user and Intra-user IC with SIC

In this case, both inter-user and intra-user interference are cancelled by SIC. The inter-user IC using SIC is the same as described in the Hybrid SIC II Algorithm, and the intra-user IC using SIC is the same as described in the Hybrid SIC I Algorithm.

Baseband Precoder The baseband precoder is the same as described in Equation (4.29). The detected $(k+1)$ to K-th MS data are cancelled after RF combining through SIC. According to Equation (4.33), the RF received signal of the k-th MS can be represented by

$$\begin{aligned}\bar{\mathbf{y}}_k = &\mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO}(:,i) s_{k,i} \\ &+ \sum_{j=1, j\neq i}^{N_s} \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO}(:,j) s_{k,j} \\ &+ \mathbf{W}_{RF,k}^H \mathbf{n}_k.\end{aligned} \tag{4.39}$$

The three successive terms of Equation (4.39) on different lines represent the i-th desired data stream for user k, the intra-user interference and the effective noise, respectively.

Baseband Combiner Different from the ZF baseband combining Equation (4.32), we use SIC to cancel intra-user interference of the k-th MS. Without loss of generality, supposing to detect the i-th data stream of the k-th MS, the baseband combiner to detect the i-th data stream of the k-th MS is $\mathbf{W}_{BB,k}^H(i,:)$, where $\mathbf{W}_{BB,k}$ is given in Equation (4.34). After the baseband combining, the detected i-th stream of the k-th MS is given by

$$\begin{aligned}\hat{s}_{k,i} &= \mathbf{W}_{BB,k}^H(i,:) \bar{\mathbf{y}}_k \\ &= \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k}^{SO}(:,i) s_{k,i} \\ &\quad + \mathbf{W}_{BB,k}^H(i,:) \mathbf{W}_{RF,k}^H \mathbf{n}_k.\end{aligned} \tag{4.40}$$

The two successive terms of the second equality of Equation (4.40) on different lines represent the i-th desired data stream for user k and the effective noise, respectively. Then we use the same procedure described in Equation (4.20) to (4.24) to detect the rest data streams of the k-th MS.

In summary, we provide inter-user IC with SIC and intra-user IC with SIC in Algorithm 8. To distinguish, we refer to Algorithm 8 as the HBF SIC III Algorithm.

Algorithm 8 HBF SIC III

Input: $\mathbf{H}_k, 1 \leq k \leq K$
Analog Stage
1. **for** $k \leq K$ **do**
2. Calculate $\mathbf{W}_{RF,k}$ according to Equation (4.12)
3. Calculate $\mathbf{H}_{int}$ according to Equation (4.11)
4. **end for**
5. Calculate $\mathbf{F}_{RF}$ according to Equation (4.13)
6. Calculate $\tilde{\mathbf{H}}_k, k = 1, 2 \cdots, K$, according to Equation (4.14)
End Analog Stage
Digital Stage
7. **for** $k \leq K$ **do**
8. **if** $k < K$ **do**
9. $\bar{\mathbf{H}}_k^{SIC} = \left[\tilde{\mathbf{H}}_{k+1}^T, \tilde{\mathbf{H}}_{k+2}^T, \cdots, \tilde{\mathbf{H}}_K^T\right]^T$
10. $\bar{\mathbf{H}}_k^{SIC} = \hat{\mathbf{U}}_k \hat{\mathbf{\Sigma}}_k \hat{\mathbf{V}}_k^H$
11. $\mathbf{F}_{BB,k}^{SO} = \hat{\mathbf{V}}_k^{\left(N_{RF}^t - (K-k)N_{RF}^r\right)} \mathbf{G}_k$
12. **else**
13. $\tilde{\mathbf{H}}_K = \hat{\mathbf{U}}_K \hat{\mathbf{\Sigma}}_K \left[\hat{\mathbf{V}}_K^{(N_{RF}^r)}, \hat{\mathbf{V}}_K^{\left(N_{RF}^t - N_{RF}^r\right)}\right]^H$
14. $\mathbf{F}_{BB,K}^{SO} = \hat{\mathbf{V}}_K^{(N_{RF}^r)} \mathbf{G}_K$
15. **end if**
16. **end for**
17. **Obtain** $\mathbf{F}_{BB}^{SO} = \left[\mathbf{F}_{BB,1}^{SO}, \cdots, \mathbf{F}_{BB,K}^{SO}\right]$
18. **for** $k \leq K$ **do**
19. $\mathbf{H}_{Equ,k} = \mathbf{W}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB}^{SO}$
20. **for** $i \leq N_s$ **do**
21. Calculate $\bar{\mathbf{H}}_{Equ,k}^i$ according to Equation (4:23)
22. $\mathbf{G}_{k,i} = \left(\left(\bar{\mathbf{H}}_{Equ,k}^i\right)^H \bar{\mathbf{H}}_{Equ,k}^i\right)^{-1} \left(\bar{\mathbf{H}}_{Equ,k}^i\right)^H$
23. $\mathbf{W}_{BB,k}(i,:) = \mathbf{G}_{k,i}\,(i,:)$
24. **end for**
25. **end for**
End Digital Stage
Output $\mathbf{F}_{RF}, \mathbf{F}_{BB}^{SO}, (\mathbf{W}_{RF,k}, \mathbf{W}_{BB,k})\,, k = 1 : K$

Detection stream Order As discussed before, the detection data stream order can be determined by post-detection SINR, which is given by Equation (4.25). Note that $\mathbf{F}_{BB,k}$ in Equation (4.25) should be updated by $\mathbf{F}_{BB,k}^{SO}$.

In this section, we have proposed three SIC-aided hybrid beamforming algorithms according to whether SIC is performed to cancel inter-user interference, intra-user interference, or both. For clarity, the IC implementation of the proposed algorithms is summarized in Table 4.1.

Table 4.1: IC implementation of the proposed algorithms

Algorithms	**Inter-user IC**	**Intra-user IC**
HBF SIC I	Baseband precoder (linear)	Baseband combiner (SIC)
HBF SIC II	Baseband precoder (linear), after RF combiner (SIC)	Baseband combiner (linear)
HBF SIC III	Baseband precoder (linear), after RF combiner (SIC)	Baseband combiner (SIC)

4.4 Computational Complexity Analysis

In this section, we analyze the complexity of implementing the proposed HBF SIC algorithms. For comparison, we also present the complexity in implementing the fully digital SVD (SVD-based single-user) [58], the fully digital BD [101], HBD [59], RF phase extraction baseband BD (PEHBD) [62], hybrid MMSE [100] and the hybrid THP beamforming algorithms [107]. In order to focus on the main ideas, we consider the simplified case where each MS is equipped with the same number of antennas N_r and RF chains N_{RF}^r. Furthermore, each MS is assumed to support the same number of data streams N_s. In Table 4.2, we enumerate the RF beamforming, digital beamforming and total complexity in undertaking major computational procedures. For example, the RF and baseband beamforming complexity of HBD algorithm is $\mathcal{O}(KN_{RF}^r N_t N_r^2)$ and $\mathcal{O}(K(K-1)^2(N_{RF}^r)^2 N_{RF}^t) + \mathcal{O}(K(N_{RF}^r)^3) \approx \mathcal{O}(K^3(N_{RF}^r)^2 N_{RF}^t)$, respectively.

For fully digital SVD beamforming, the total complexity is dominated by SVD of the overall channel matrix with dimension $KN_r \times N_t$, yielding the complexity order of $\mathcal{O}(K^2 N_t N_r^2)$. For fully digital BD beamforming, the total complexity is dominated by K times of SVDs of $\bar{\mathbf{H}}_k$, yielding the complexity order of $\mathcal{O}(K(K-1)^2 N_t N_r^2)$. For HBD, the total complexity is dominated by RF beamforming, and therefore can be shown to be the complexity order of $\mathcal{O}(KN_{RF}^r N_t N_r^2)$. For the hybrid THP algorithm, which starts by constructing the fully digital THP precoder and combiner, and then

performs K times of SVDs in the BD-GMD, it can be shown to be of complexity $\mathcal{O}(KN_sN_{RF}^tN_t^2) + \mathcal{O}(KN_r^2N_t)$ [107]. For all other hybrid beamforming algorithms, the total complexity is dominated by K times of SVDs of $\mathbf{H}_k$, yielding the same complexity order of $\mathcal{O}(KN_r^2N_t)$. Therefore, the total complexity of the proposed SIC-aided hybrid beamforming algorithms are comparable to PEHBD, hybrid MMSE and the hybrid THP algorithms.

When considering the computational complexity of baseband beamforming part of the hybrid schemes, the required extra complexity of the HBF SIC I algorithm compared to the existing linear algorithms is dominated by KN_s times of matrix inverse of $\hat{\mathbf{H}}_k$ when performing SIC to cancel intra-user interference, and therefore can be shown to be of complexity $\mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)+\mathcal{O}(KN_{RF}^rN_s^3) \approx \mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)$. Furthermore, the complexity of the HBF SIC II algorithm is dominated by $K-1$ times of SVDs when performing SIC to cancel inter-user interference, yielding the complexity order of $\mathcal{O}(\frac{1}{3}K^3(N_{RF}^r)^2N_{RF}^t)$, K times of SVDs with the complexity order of $\mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)$ and K times of matrix inverse with the complexity order of $\mathcal{O}(K(N_{RF}^r)^2N_{RF}^t)$ when maximizing the spectral efficiency of each MS. Therefore the complexity order of the HBF SIC II algorithm is simplified to $\mathcal{O}(\frac{4}{3}K^3(N_{RF}^r)^2N_{RF}^t)$. Similarly, the complexity order of the HBF SIC III algorithm is $\mathcal{O}(\frac{4}{3}K^3(N_{RF}^r)^2N_{RF}^t)$. Moreover, the complexity order of the Hybrid MMSE algorithm is $\mathcal{O}(K^3(N_{RF}^r)^3N_{RF}^t)$ [100]. Therefore, the baseband beamforming complexity of fully digital BD $\gg$ fully digital $\gg$ hybrid THP $\gg$ hybrid MMSE $\gg$ HBF SIC III = HBF SIC II $>$ HBF SIC I = PFHBD = HBD. It should be noted that the complexity of the HBF SIC III algorithm is a little higher than the HBF SIC II algorithm when considering all procedures since the HBF SIC III algorithm uses SIC to cancel intra-user interference while the HBF SIC II algorithm uses ZF to cancel intra-user interference.

Table 4.2: Comparison of computational complexity

Algorithms	**Calculate $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$**	**Calculate $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$**	**Total Complexity**
Fullly Digital SVD	Not Involved	$\mathcal{O}(K^2N_tN_r^2)$	$\mathcal{O}(K^2N_tN_r^2)$
Fullly Digital BD	Not Involved	$\mathcal{O}(K(K-1)^2N_tN_r^2)$	$\mathcal{O}(K(K-1)^2N_tN_r^2)$
HBD	$\mathcal{O}(KN_{RF}^rN_tN_r^2)$	$\mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_{RF}^rN_tN_r^2)$
PFHBD	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_r^2N_t)$
Hybrid MMSE	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(K^3(N_{RF}^r)^3N_{RF}^t)$	$\mathcal{O}(KN_r^2N_t)$
HBF SIC III	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(\frac{4}{3}K^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_r^2N_t)$
HBF SIC II	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(\frac{4}{3}K^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_r^2N_t)$
HBF SIC I	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(K^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_r^2N_t)$
Hybrid THP	$\mathcal{O}(KN_sN_{RF}^tN_t^2)$	$\mathcal{O}(KN_r^2N_t)$	$\mathcal{O}(KN_sN_{RF}^tN_t^2)+\mathcal{O}(KN_r^2N_t)$

4.5 Simulation Results

In this section, simulation results are presented to illustrate the sum SE performance of the proposed algorithms as well as their performance robustness with finite resolution phase shifters and imperfect CSI. In view of hardware cost and power consumption of RF chains, the least number of RF chains are considered at transceivers.

Unless stated otherwise, the system parameters are set as follows. Both BS and MS are equipped with ULA with the number of antennas $N_t = 256$ and $N_r = 16$. For simplicity, we suppose that the channels from BS to each MS are with the same number of scattering clusters $N_c = 8$. The AoA and AoD of the clusters are assumed to be uniformly distributed in $[0, 2\pi]$. The signal-to-noise ratio is defined as SNR $= P/\sigma_n^2$. All the results are simulated over 3000 channel realizations.

4.5.1 Sum Spectral Efficiency Evaluation

In this set of simulation results, we illustrate the sum SE performance of the proposed SIC-aided HBF algorithms in different configurations. The state of the art mmWave multi-user massive MIMO systems for comparison include HBD [59], PEHBD [62], Hybrid MMSE [100] and fully digital BD scheme [101]. The benchmark is provided by the sum of the maximum achievable spectral efficiency of the single-user fully digital SVD beamforming, which is implemented by transmitting the same number of data streams over the same channels in the absence of inter-user interference. Specifically, we stack the channel of each MS $\mathbf{H}_k$ into an overall channel matrix $\mathbf{H}$ and assume the absence of inter-user interference. The full digital precoding and combining matrix are composed of KN_s right singular vectors and KN_s left singular vectors corresponding to the KN_s largest singular values of $\mathbf{H}$, respectively.

Figure 4.2 to Figure 4.4 illustrate the sum spectral efficiency achieved by the proposed SIC-aided hybrid beamforming algorithms in contrast with fully digital beamforming and traditional hybrid beamforming designs considering three different scenarios. In the simulation of Figure 4.2, $N_{RF}^t = 16$, $K = 8$, $N_{RF}^r = N_s = 2$, the inter-user interference is dominant in this scenario. In the simulation of Figure 4.3, $N_{RF}^t = 16$, $K = 4$, $N_{RF}^r = N_s = 4$, both the inter-user interference and intra-user interference are dominant in this scenario. In the simulation of Figure 4.4, $N_{RF}^t = 8$, $K = 2$, $N_{RF}^r = N_s = 4$, the intra-user interference is dominant in this scenario.

Figure 4.2 to Figure 4.4, we can observe that the proposed HBF SIC algorithms achieve slightly lower sum spectral efficiency compared to the benchmark in the case

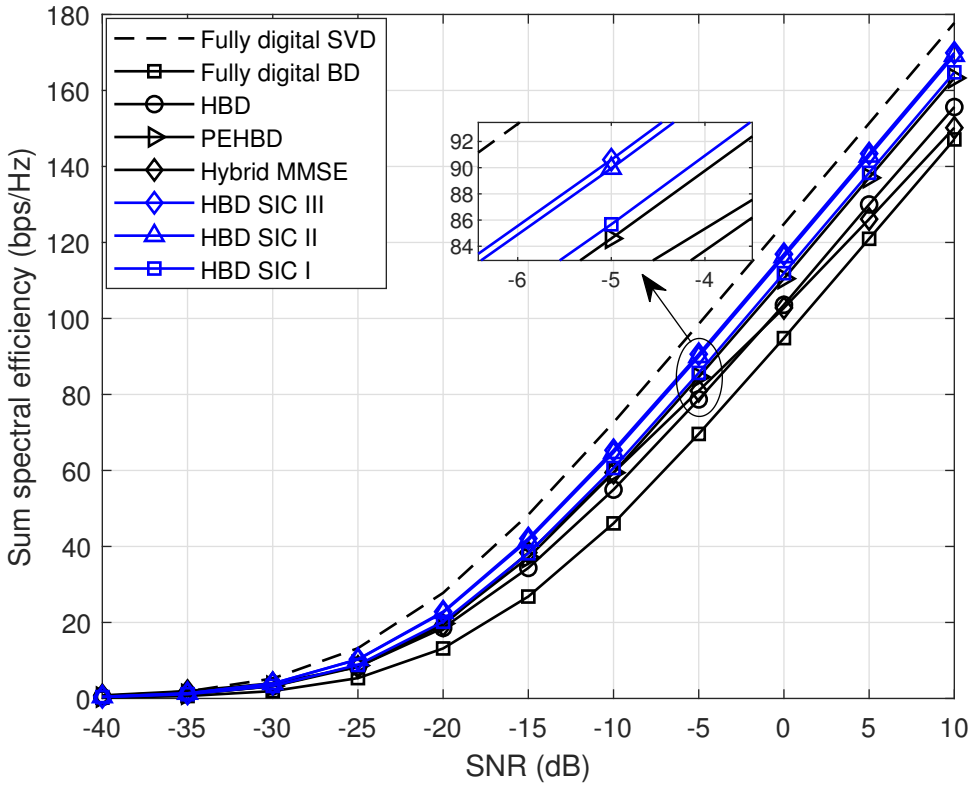

Figure 4.2: Sum SE performance comparison of different algorithms with $K = 8$, $N_{\mathrm{s}} = 2$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 2$.

of 8 MSs and perform very close to the benchmark in the case of 4 and 2 MSs. Furthermore, with the more powerful nonlinear beamforming, all the proposed HBF SIC algorithms exhibit substantially higher sum spectral efficiency compared to the traditional linear hybrid beamforming algorithms. Finally, it can be observed from Figure 4.4 that the HBF SIC II algorithm shows lower sum spectral efficiency than the HBF SIC I algorithm. This is because in the simulation set up of Figure 4.4, the intra-user interference is dominant. The HBF SIC II algorithm uses ZF to cancel intra-user interference while the HBF SIC I algorithm uses SIC to cancel intra-user interference. Therefore, the HBF SIC II algorithm shows lower sum spectral efficiency than the HBF SIC I algorithm.

Figure 4.5 and Figure 4.6 further evaluate the sum spectral efficiency at SNR = 0 dB when varying the number of BS antennas and MS antennas with $N_{RF}^{t} = 16$, $K = 8$, $N_{RF}^{r} = N_s = 2$. In the simulation of Figure 4.5, $N_r = 16$, the number of antennas at BS changes from $N_t = 20$ to $N_t = 500$. In the simulation of Figure

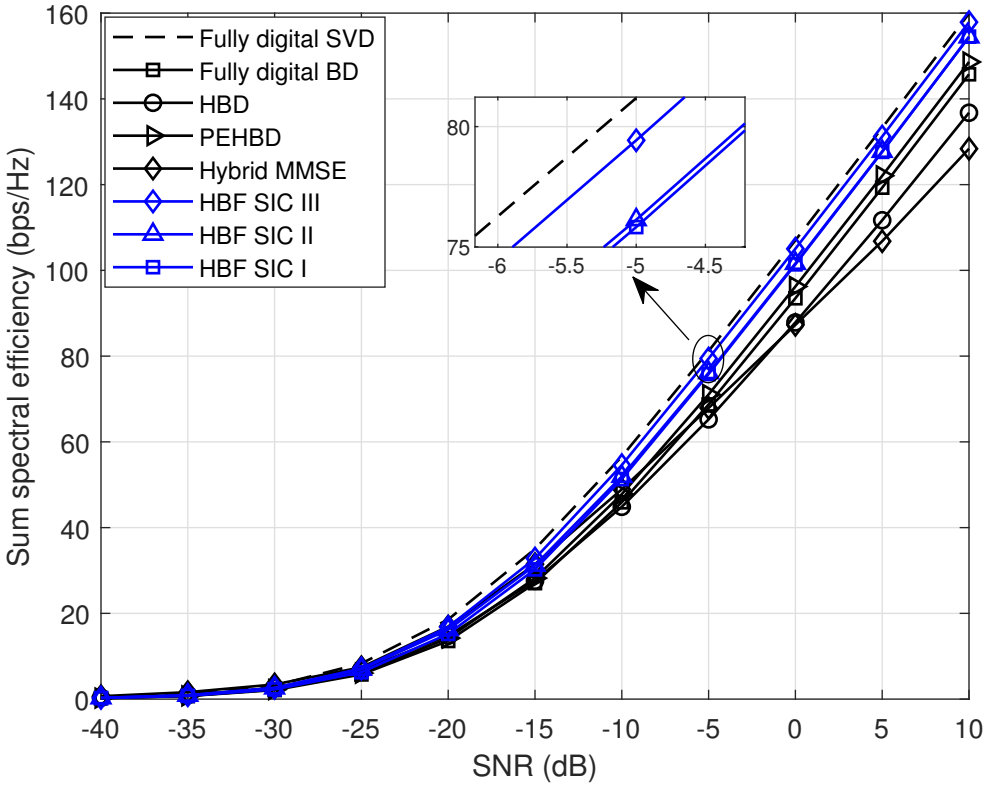

Figure 4.3: Sum SE performance comparison of different algorithms with $K = 4$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

4.6, $N_t = 256$, the number of antennas at MS changes from $N_r = 10$ to $N_r = 100$. The figures show a considerable performance gain by the proposed HBF SIC I to III algorithms over the PEHBD, HBD and Hybrid MMSE algorithms, especially for the proposed HBF SIC II and III algorithms. The reason is the inter-user interference is the main interference in this scenario, and it is cancelled by SIC in HBF SIC II and III algorithms, which boosts the sum spectral efficiency performance.

With the same system configuration as that of Figure 4.2, and K increases from 2 to 16, $N_{RF}^t = KN_{RF}^r = KN_s$ in Figure 4.7. In this scenario, the sum spectral efficiency of the proposed SIC-aided hybrid beamforming, PEHBD and HBD algorithms increase as the number of MSs K grows. A considerable performance improvement is observed by the proposed algorithms over the ones in the literature, especially for the proposed HBF SIC II and III. The reason is that inter-user interference becomes the main interference when K increases, and it is cancelled by SIC in HBF SIC II and III algorithms, which boosts the sum spectral efficiency performance. As for the

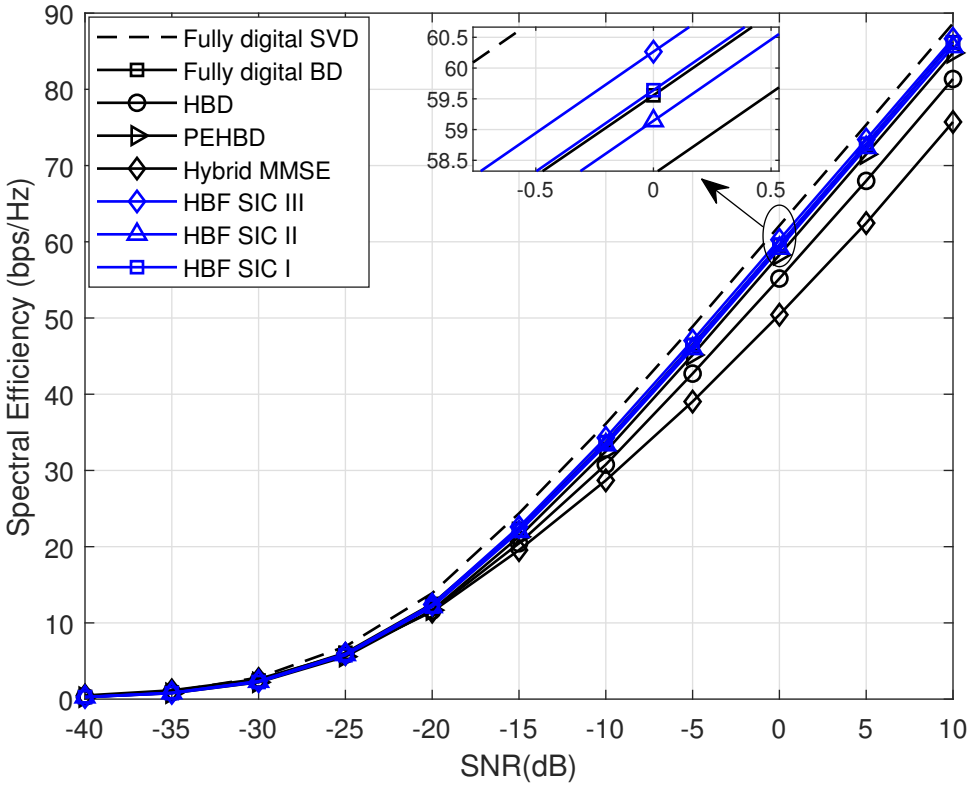

Figure 4.4: Sum SE performance comparison of different algorithms $K = 2$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

fully digital BD algorithm, it reaches a peak sum spectral efficiency at a certain K, and then the sum spectral efficiency will decrease along with increasing K beyond an optimal value. This is because when K grows beyond an optimal value, the inter-user interference becomes substantially severe, which degrades the sum spectral efficiency performance.

Figure 4.8 compares the sum spectral efficiency achieved by the proposed HBF SIC algorithms and the Hybrid THP algorithm [107]. The system configuration is the same as that of Figure 4.2. It can be observed from Figure 4.8 that the proposed HBF SIC III and HBF SIC II are superior to Hybrid THP while Hybrid THP is superior to the proposed HBF SIC I. The reason is that Hybrid THP cancels the inter-user interference by the BD algorithm and suppresses the intra-user interference by SIC at the transmitter end. While the proposed HBF SIC III and HBF SIC II algorithms cancel the inter-user interference by SIC, and therefore, the proposed HBF SIC III and HBF SIC II are superior to the Hybrid THP algorithm.

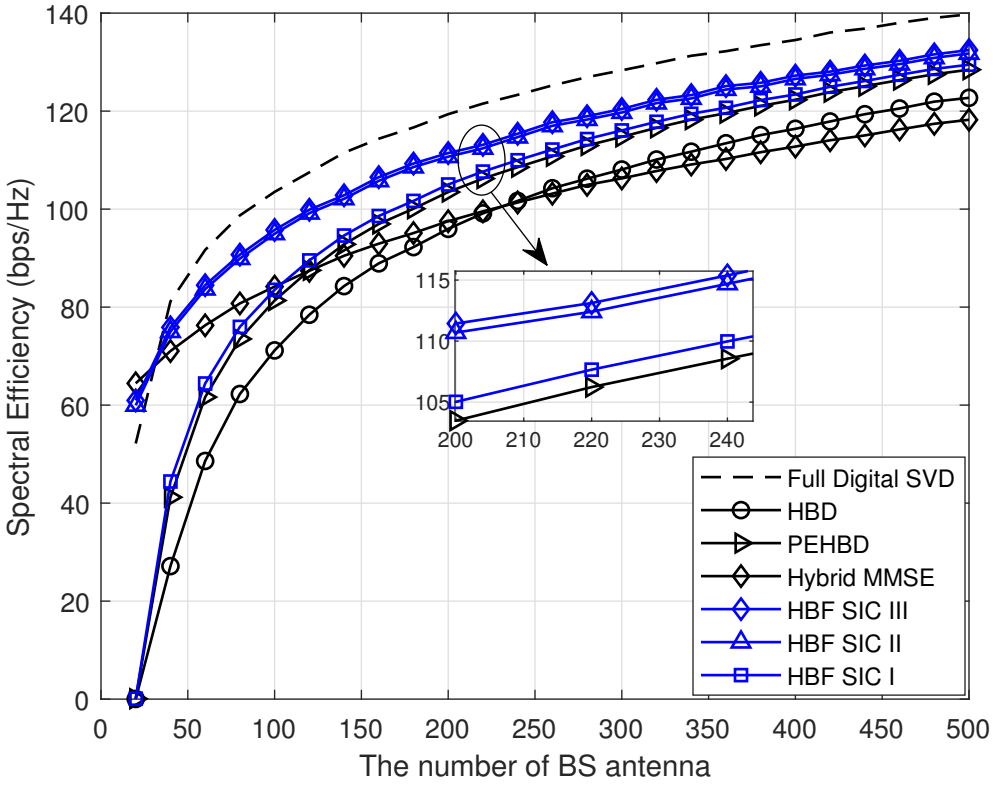

Figure 4.5: Sum SE performance with different number BS antenna.

Figure 4.9 illustrates the sum spectral efficiency improvement achieved by the proposed ordered HBF SIC algorithms with the same system configuration as that of Figure 4.2. In Figure 4.9, when SNR = -5 dB, we can observe the sum spectral efficiency improvement achieved by the proposed HBF SIC III algorithm with optimal data stream order (determined by Equation (4.25)) is about 0.6% compared to the proposed HBF SIC III algorithm with arbitrary data stream order. The sum spectral efficiency improvement achieved by the proposed HBF SIC I algorithm with optimal data stream order (determined by Equation (4.25)) is about 0.8% compared to the proposed HBF SIC I algorithm with arbitrary data stream order.

4.5.2 Robustness Evaluation

In this set of simulation results, we illustrate the robustness of the proposed SIC-aided HBF algorithms with finite resolution phase shifters and imperfect CSI.

Figure 4.10 illustrates the sum spectral efficiency achieved by the proposed SIC-

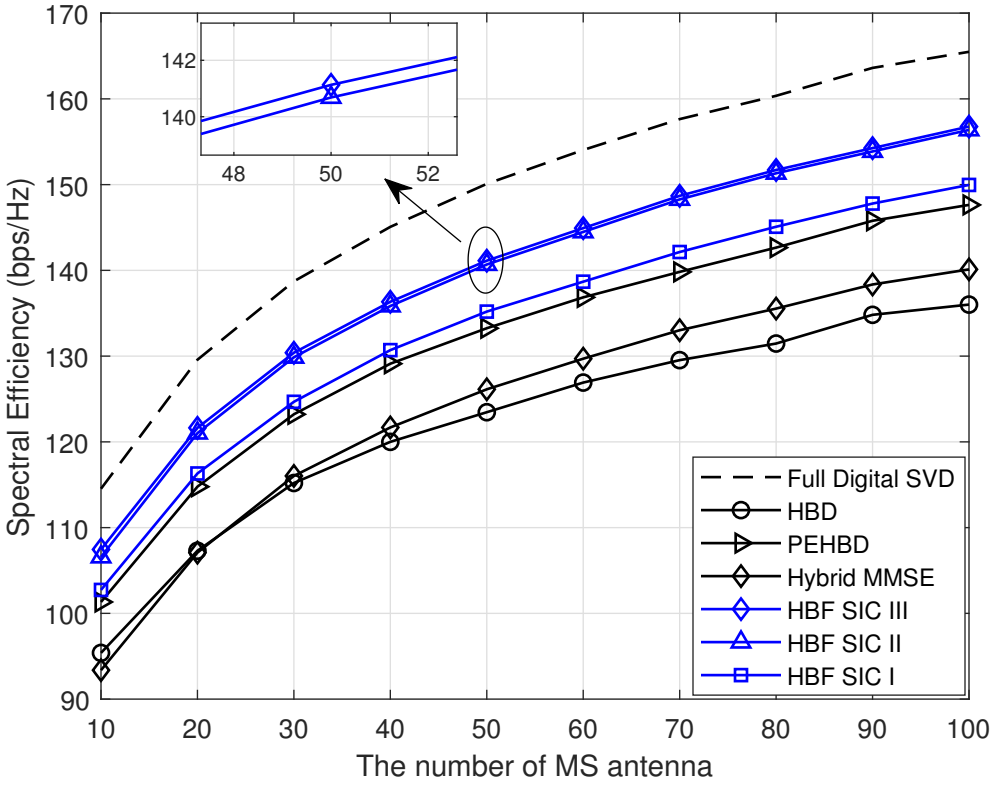

Figure 4.6: Sum SE performance with different number MS antenna.

aided hybrid beamforming algorithms with finite resolution phase shifters. The system configuration is the same as that of Figure 4.2. As observed from Fig. 4.10, when SNR $= -5$ dB, the performance of the proposed HBF SIC algorithms for 2-bit phase shifters (B $= 2$ bits is used in the figures) and 1-bit phase shifters (B $= 1$ bit is used in the figures) can achieve over 91% and 64% performance of infinite resolution phase shifters (B $= \infty$ is used in the figures), respectively.

Figure 4.11 evaluates the impact of imperfect CSI on the proposed SIC-aided hybrid beamforming algorithms with the same system configuration as that of Figure 4.2. The estimated channel matrix can be modelled as Equation (3.10). As shown in Figure 4.11, three proposed HBF SIC algorithms experience similar degradation in the presence of channel estimation error. The performance gap to the perfect CSI is less than 1 dB when $\xi = 0.9$, and about 2.5 dB when $\xi = 0.7$, for the considered imperfect CSI model.

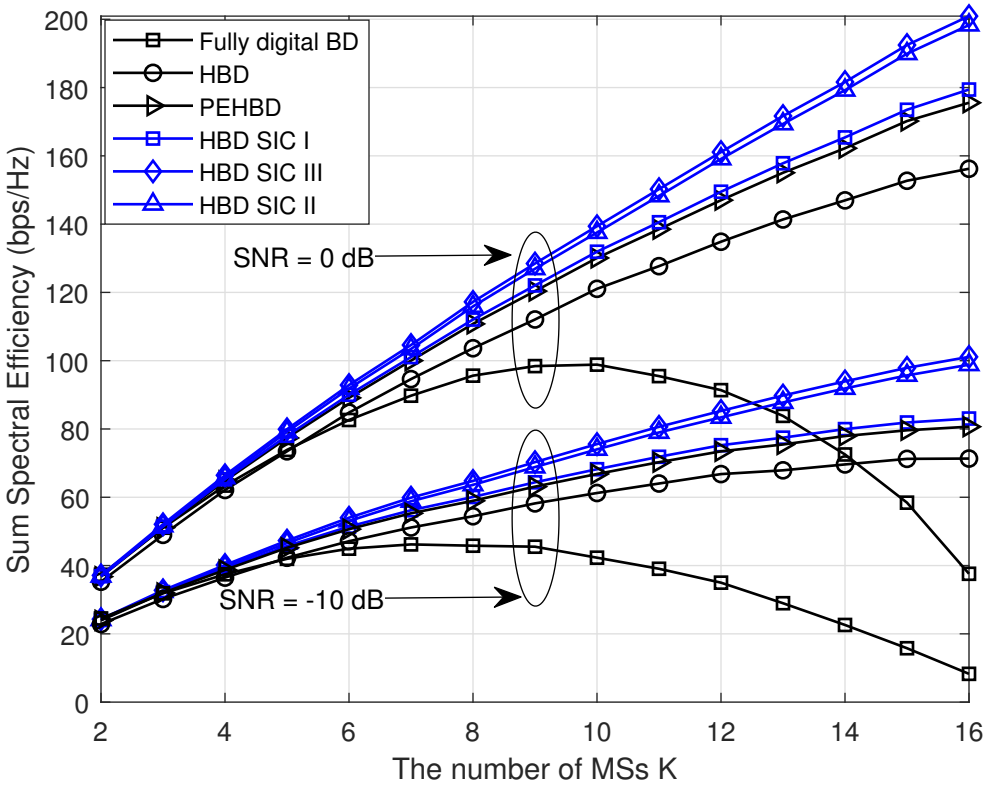

Figure 4.7: Sum SE performance with different number of MS.

4.6 Summary

In this chapter, we have investigated the non-linear hybrid beamforming design with interference cancellation for downlink mmWave multi-user massive MIMO systems. Based on the proposed framework, three SIC-aided hybrid beamforming algorithms have been proposed according to the dominance of inter-user interference, intra-user interference, or both. It has been demonstrated that the proposed hybrid beamforming algorithms exhibit a comparable SE to the fully digital beamforming and outperform their existing linear counterparts at the cost of computational complexity for the SIC procedure. Furthermore, the proposed hybrid beamforming algorithms with 2-bit finite resolution phase shifters can achieve over 91% SE of infinite resolution phase shifters. Moreover, when considering channel estimation error, the performance gap to the perfect CSI is less than 1 dB when $\xi = 0.9$. Therefore, our proposed hybrid beamforming algorithms reveal the profound practically to be employed in mmWave communications.

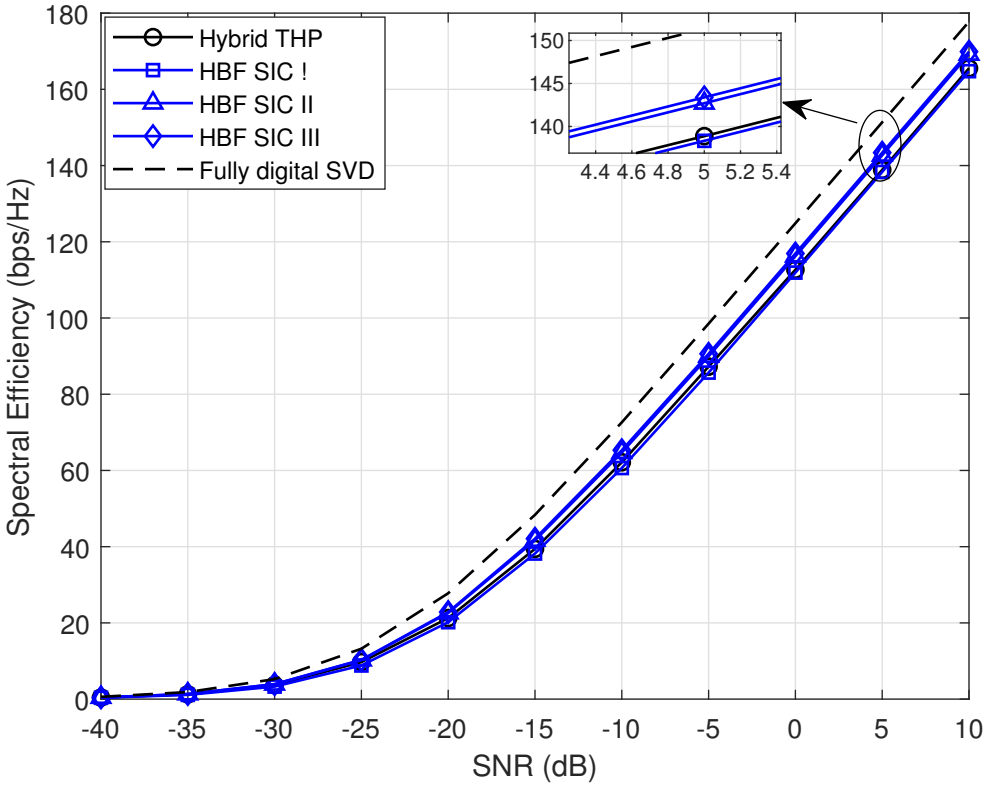

Figure 4.8: Sum SE comparison to the Hybrid THP.

It is worth mentioning that the proposed algorithms are designed for flat fading mmWave channels. In the next chapter, we will develop efficient hybrid precoding and combining algorithms for for wideband mmWave channels, which may emerge more frequently in future applications due to large bandwidth of mmWave bands.

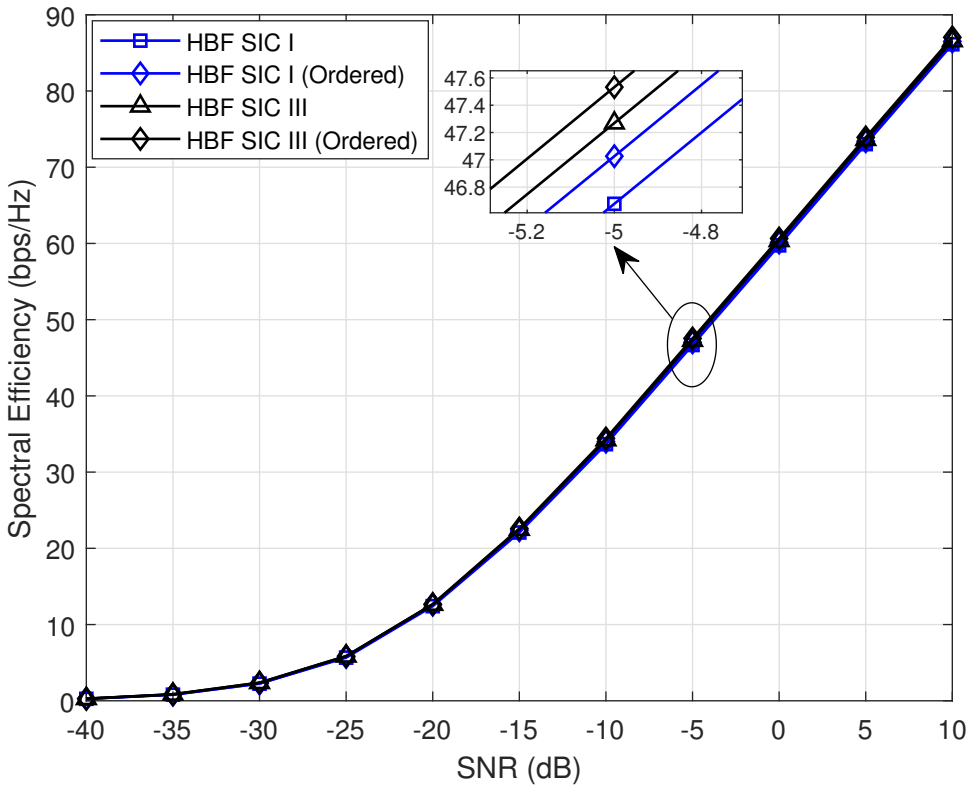

Figure 4.9: Sum SE performance comparison of proposed ordered algorithms.

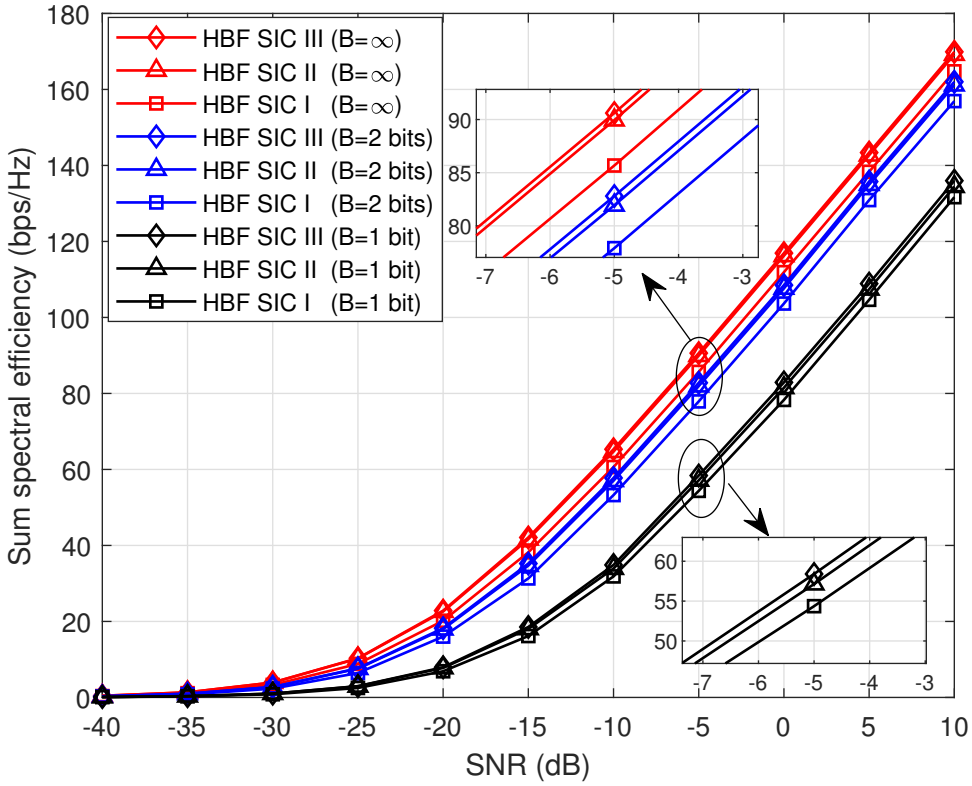

Figure 4.10: Sum SE performance with finite resolution phase shifters.

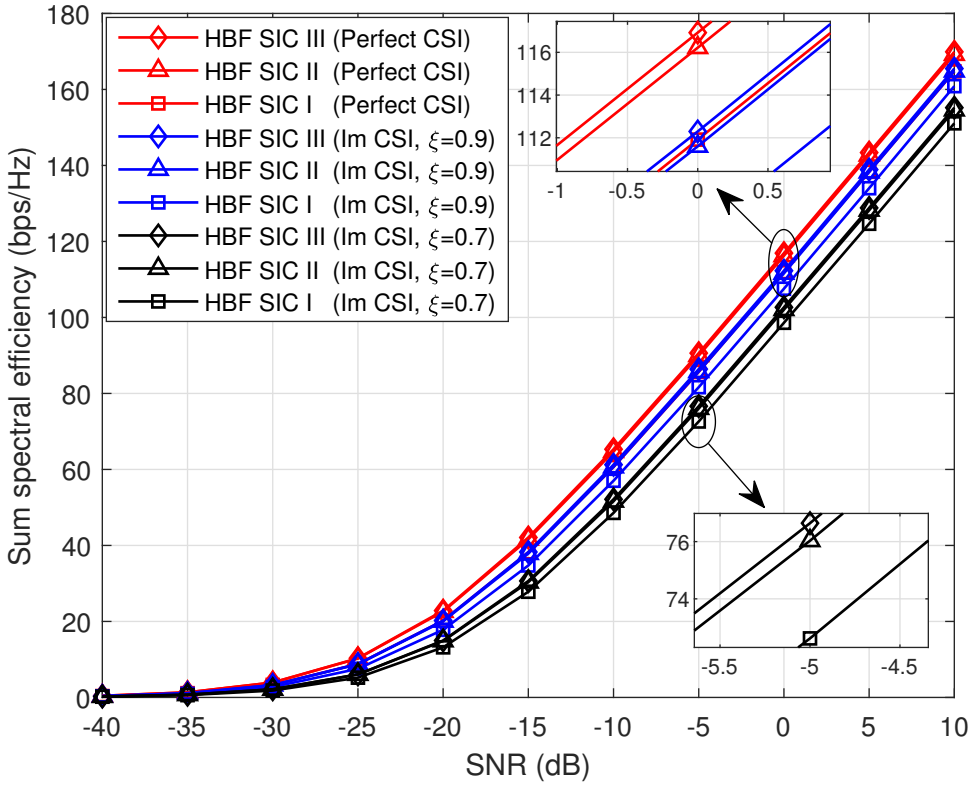

Figure 4.11: Sum SE performance with imperfect CSI.

Chapter 5

Hybrid Beamforming for Wideband Millimeter Wave Massive MIMO Systems

5.1 Introduction

In Chapter 3 and 4, we consider hybrid beamforming design for narrowband mmWave communication systems. However, mmWave communication systems will likely operate on wideband channels with frequency selective fading due to a large bandwidth in mmWave bands [72], [82]. How to design a common RF beamformer across all bands becomes the main challenge. Furthermore, the common RF beamformer in wideband channels leads to the need of more effective baseband beamforming design. In this chapter, we will focus on the development of efficient hybrid precoding and combining algorithms for for wideband mmWave channels.

5.1.1 Prior Work

The design of hybrid precoding schemes for the wideband mmWave MIMO system has attracted more interest, where typically orthogonal frequency-division multiplexing (OFDM) is utilized to convert broadband frequency-selective fading channels into multiple parallel narrowband frequency-flat fading channels. However, hybrid precoding design for narrowband scenarios cannot be directly extended to wideband scenarios due to the RF precoding and combining identical to all subcarriers, which is the main challenge in designing the hybrid precoding and combining in wideband

OFDM systems. Because when a common analog beamformer is applied over all subcarriers, the directional response of the beamformer will change with frequency. In [120], hybrid beamforming with only a single-stream transmission over MIMO-OFDM systems was considered. The solution proposed in [120] relied on the joint exhaustive search over both RF and baseband codebooks without giving specific criteria for the design of these codebooks. In [85], the authors first developed a hybrid analog-digital codebook design scheme for broadband mmWave single-user MIMO (SU-MIMO) systems, then proposed a hybrid precoding algorithm for the given codebook based on Gram-Schmidt orthogonalization. However, the authors in [85] only considered SU-MIMO systems. Furthermore, they did not specify the hybrid combiner design. In [121], the authors proposed a joint hybrid precoder and combiner design for wideband mmWave MIMO-OFDM systems with dynamic hybrid beamforming architecture. However, only point-to-point mmWave MIMO-OFDM systems are considered. In [122], the authors proposed two hybrid beamforming approaches, based on the virtual sub-array and the true-time-delay lines, respectively, to eliminate the impact of beam squint in wideband terahertz communications. However, only single-user MIMO-OFDM systems are considered. Furthermore, the digital precoder design in [122] adopts linear algorithms to cancel interference, which is insufficient due to the common RF beamformer in wideband scenarios. In [86], the authors proposed a closed-form solution to hybrid precoder for fully connected and partially connected wideband mmWave SU-MIMO systems by exploiting the covariance matrix of frequency domain channels. However, [86] assumed the fully-digital combining at the receiver. The average of the covariance matrices of frequency domain channels is also used to design hybrid precoder for wideband mmWave SU-MIMO OFDM and MU-MISO OFDM systems in [96]. However, the digital precoder design for SU-MIMO and MU-MISO in [96] adopt linear algorithms to cancel interference, which is not sufficient due to the common RF beamformer in wideband scenarios. Using the minimum mean square error (MMSE) criterion, the authors in [94] proposed a hybrid beamforming design via eigenvalue decomposition and OMP for broadband mmWave SU-MIMO systems. In [123], the authors proposed a principal component analysis (PCA)-based hybrid precoder and combiner design for broadband mmWave SU-MIMO systems. Only a small number of papers considered the design of hybrid precoding for wideband multi-user MIMO (MU-MIMO) scenarios [57, 124–126]. In [124], the authors proposed a hybrid beamforming for wideband mmWave MU-MIMO OFDM systems. In particular, the analog precoder and combiner design are based on tensor unfold-

ing, and the digital precoder is designed to maximize the signal-to-leakage-plus-noise ratio (SLNR). The authors in [57] established the required number of RF chains and phase-shifters such that hybrid beamforming achieves the same performance as that of the digital beamforming. The work in [57], however, is limited to the case of single-antenna users. In [125], the authors proposed to design the precoders and combiners for MU-MIMO OFDM systems by leveraging the duality between the downlink and uplink. However, the authors neglected the inter-user interference in the design of the uplink precoders. Furthermore, the computational complexity is very high due to multiple iterations. In [126], the authors proposed a non-uniformly spaced quantization codebook based RF precoding without considering the inter-user interference and block diagonalization (BD) based baseband precoding algorithm to cancel the inter-user interference.

As mmWave communication systems are expected to be deployed on wideband channels with frequency selective fading, how to design a common RF beamformer across all subcarriers becomes the main challenge. Furthermore, the common RF beamformer in wideband channels leads to the need of more effective baseband schemes. However, most of the existing works are focused on linear algorithms to cancel inter-user and intra-user interference. These two important features motivate us to consider the hybrid beamforming design for wideband mmWave MU-MIMO OFDM systems. In this chapter, we develop hybrid beamforming solutions for wideband mmWave MU-MIMO OFDM systems. The analog beamforming aiming to maximize the mutual information or spectral efficiency is frequency flat, while the digital beamforming is different for each subcarrier aiming to cancel inter-user interference and intra-user interference. The rest of this chapter is organized as follows. In Section 5.2, we introduce wideband mmWave MU-MIMO hybrid beamforming system model, wideband mmWave channel model and the formulated optimization problem. Then, the proposed hybrid precoding and combining algorithm are presented in Section 5.3. In Section 5.4, the effect of channel estimation error on the proposed algorithm is analyzed. The computational complexity of the proposed algorithm is analyzed in Section 5.5. The simulation results used to evaluate the spectral efficiency performance of the proposed algorithms are provided in Section 5.6, followed by summary in Section 5.7. In this chapter, k refers to the kth subcarrier.

5.2 System Model and Problem Formulation

In this section, we introduce the wideband mmWave MU-MIMO hybrid beamforming structure and wideband mmWave channel model.

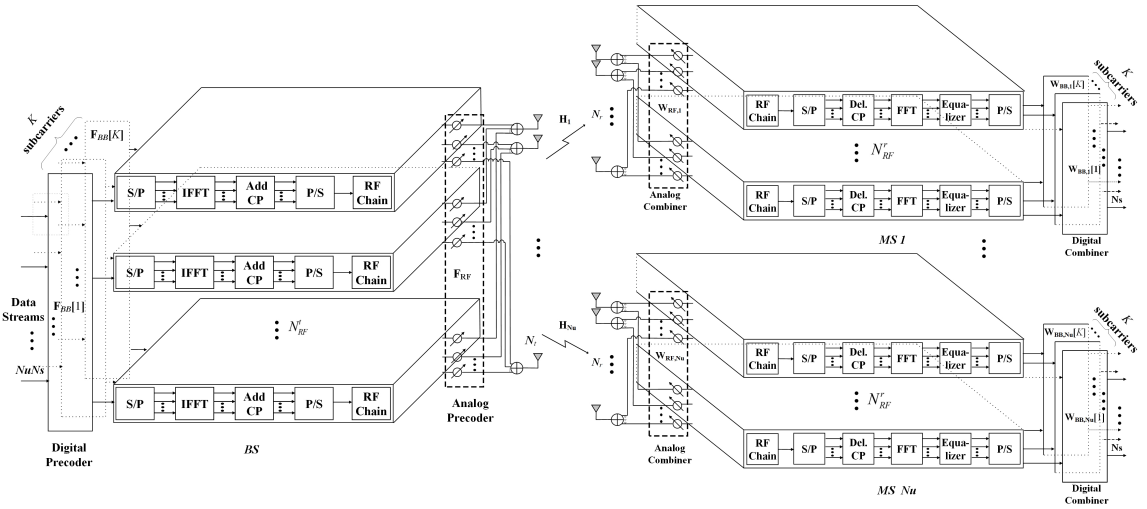

Figure 5.1: System block diagram of a fully connected downlink multi-user massive MIMO with hybrid precoding/combining structure.

5.2.1 System Model

We consider the OFDM based wideband mmWave MU-MIMO hybrid beamforming communication system as shown in Figure 5.1 where a BS with N_t massive antennas and N_{RF}^t limited RF chains is simultaneously communicating with N_u MSs. Each MS is equipped with N_r antennas and N_{RF}^r RF chains. In general, the number of data streams can be different for different MSs. However, for simplicity, this paper restricts attention to the scenario with an equal number of data streams for all users. Without loss of generality, we assume each MS support N_s ($N_s \leq N_{RF}^r$) data streams, which means total $N_u N_s$ ($N_u N_s \leq N_{RF}^t$) data streams are transmitted by the BS. To guarantee the effectiveness of the hybrid processing structure, the number of RF chains is constrained by $N_u N_s \leq N_{RF}^t \ll N_t$ for the BS and $N_s \leq N_{RF}^r \ll N_r$ for each MS. Note that to support N_s data streams transmission for each user, the least number of RF chains at the BS and each MS are $N_{RF}^t = N_u N_s$ and $N_{RF}^r = N_s$, respectively.

At the BS, the $N_u N_s$ data streams $\mathbf{S}[k]$ at each subcarrier k are first precoded by an $N_{RF}^t \times N_u N_s$ baseband precoder $\mathbf{F}_{BB}[k] = [\mathbf{F}_{BB,1}[k], \mathbf{F}_{BB,2}[k], \cdots, \mathbf{F}_{BB,N_u}[k]]$,

where $\mathbf{F}_{BB,i}[k] \in \mathbb{C}^{N_{RF}^t \times N_s}$. After serial to parallel (S/P) transformation, the data streams are transformed to the time domain using N_{RF}^t K-point inverse fast Fourier Transform (IFFT). Note that it is assumed all subcarriers are used in our system model, therefore, the data stream length is equal to the number of subcarriers. After adding a cyclic prefix (CP) of length L_{cp} and parallel to serial (P/S) transformation, the data streams are finally precoded by an $N_t \times N_{RF}^t$ analog precoder $\mathbf{F}_{RF}$, which is implemented through an array of analog phase shifters. Since the analog precoder is a post-IFFT module, it is important to emphasize here that the analog precoder is identical for all subcarriers, which means that the analog precoder is frequency flat while the digital precoder is frequency selective. This is the main distinguishing feature of OFDM-based wideband hybrid beamforming compared to the narrowband scenario. Since the phase shifters in the analog RF precoder can only change the phase of the transmitted signal, each entry of $\mathbf{F}_{RF}$ is of constant modulus. We normalize its entries to satisfy $|\mathbf{F}_{RF}(i,j)| = 1/\sqrt{N_t}$. Furthermore, to meet the total transmit power constraint for each subcarrier, $\mathbf{F}_{BB}[k]$ is normalized to satisfy $\|\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\|_F^2 = N_u N_s$. The discrete-time transmitted signal at subcarrier k can therefore be modeled as

$$\mathbf{X}[k] = \mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{S}[k] = \sum_{i=1}^{N_u} \mathbf{F}_{RF}\mathbf{F}_{BB,i}[k]\mathbf{s}_i[k], \tag{5.1}$$

where $\mathbf{S}[k] = \left[\mathbf{s}_1^T[k], \mathbf{s}_2^T[k], \cdots, \mathbf{s}_{N_u}^T[k]\right]^T \in \mathbb{C}^{N_u N_s \times 1}$ is the signal vector for total transmitted symbols of all N_u users, and $\mathbf{s}_i[k] \in \mathbb{C}^{N_s \times 1}$ denotes the transmitted symbol of user i at subcarrier k with $\mathbb{E}\left\{\mathbf{s}_i[k]\mathbf{s}_i^H[k]\right\} = P_i\mathbf{I}_{N_s}$. Assuming a block fading channel model, the received signal of the i-th MS at subcarrier k is given by

$$\begin{aligned}\mathbf{r}_i[k] &= \mathbf{H}_i[k]\mathbf{X}[k] + \mathbf{n}_i[k] \\ &= \mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{S}[k] + \mathbf{n}_i[k],\end{aligned} \tag{5.2}$$

where $\mathbf{H}_i[k] \in \mathbb{C}^{N_r \times N_t}$ represents the frequency selective fading channel between MS i and the BS at subcarrier k, $\mathbf{n}_i[k] \in {}^{N_r \times 1}$ is the additive complex Gaussian noise and each entry of $\mathbf{n}_i[k]$ follows the independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and variance σ^2.

At the receiver i, assuming perfect time and frequency synchronization, the received signal of all subcarriers are initially processed by an analog combiner, $\mathbf{W}_{RF,i} \in \mathbb{C}^{N_r \times N_{RF}^r}$, which is implemented by analog phase shifters similar to the BS. We also normalize its entries to satisfy $|\mathbf{W}_{RF}(i,j)| = 1/\sqrt{N_r}$. After S/P transformation, the

CP is removed and N_{RF}^r K-point fast Fourier Transforms (FFTs) are applied to transform the data streams to the frequency domain. Then N_{RF}^r frequency domain equalizers are applied to recover the frequency domain signals. After P/S transformation, finally the received signal is processed by a digital combiner $\mathbf{W}_{BB,i}[k] \in \mathbb{C}^{N_{RF}^r \times N_s}$, the receiver obtains the final processed signal given by

$$\begin{aligned}\mathbf{y}_i[k] &= \mathbf{W}_{BB,i}^H[k]\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k]\mathbf{s}_i[k] \\ &+ \mathbf{W}_{BB,i}^H[k]\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\sum_{j=1,j\neq i}^{N_u}\mathbf{F}_{RF}\mathbf{F}_{BB,j}[k]\mathbf{s}_j[k] \\ &+ \mathbf{W}_{BB,i}^H[k]\mathbf{W}_{RF,i}^H\mathbf{n}_i[k].\end{aligned} \tag{5.3}$$

The three successive terms of Equation (5.3) denote the desired i-th MS signal, inter-user interference and noise, respectively. It should be noted the desired i-th MS signal, $\mathbf{y}_{de,i}[k]$ consists of the desired m-th data stream and the intra-user interference, as given by Equation (5.4).

$$\begin{aligned}\mathbf{y}_{de,i}[k] &= \mathbf{W}_{BB,i}^H[k]\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k]\mathbf{s}_i[k] \\ &= \mathbf{W}_{BB,i}^H[k](m,:)\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k](:,m)s_{i,m}[k] \\ &+ \mathbf{W}_{BB,i}^H[k](m,:)\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\sum_{n=1,n\neq m}^{N_s}\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k](:,n)s_{i,n}[k].\end{aligned} \tag{5.4}$$

It can be seen from Equation (5.3) and Equation (5.4) that the received signal of MS i at subcarrier k includes the desired data stream, the intra-user interference (or inter-stream interference of the i-th MS), the inter-user interference from other $N_u - 1$ MSs and the effective noise. When the Gaussian symbol is utilized by the BS, the spectral efficiency achieved by the n-th data stream for the i-th MS at subcarrier k will be

$$R_{i,n}[k] = \log_2\left(|1 + SINR_{i,n}[k]|\right). \tag{5.5}$$

The sum spectral efficiency of all MSs at subcarrier k is given by

$$R[k] = \sum_{i=1}^{N_u}\sum_{n=1}^{N_s}R_{i,n}[k] = \sum_{i=1}^{N_u}\sum_{n=1}^{N_s}\log_2\left(|1 + SINR_{i,n}[k]|\right), \tag{5.6}$$

where $SINR_{i,n}[k] = \frac{P_s[k]}{P_I[k]+P_N[k]}$ denotes SINR of the n-th stream of user i at subcarrier

k. $P_s[k]$, $P_I[k]$ and $P_N[k]$ are desired signal power, interference power and noise power, given by Equation (5.7), (5.8) and (5.9), respectively.

$$P_s[k] = \frac{P_i}{KN_s}\mathbf{W}_{BB,i}^H[k](n,:)\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k](:,n)\mathbf{F}_{BB,i}{}^H[k](:,n)\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:). \tag{5.7}$$

$$\begin{aligned} P_I[k] = {} & \tfrac{P_i}{KN_s}\mathbf{W}_{BB,i}^H[k](n,:)\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\times \\ & \sum_{m=1,m\neq n}^{N_s}\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k](:,m)\mathbf{F}_{BB,i}{}^H[k](:,m)\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:) \\ & + \tfrac{1}{KN_s}\mathbf{W}_{BB,i}^H[k](n,:)\mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\times \\ & \sum_{j=1,j\neq i}^{N_u}P_j\mathbf{F}_{RF}\mathbf{F}_{BB,j}[k]\mathbf{F}_{BB,j}^H[k]\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:). \end{aligned} \tag{5.8}$$

$$P_N[k] = \sigma^2\mathbf{W}_{BB,i}^H[k](n,:)\mathbf{W}_{RF,i}^H\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:). \tag{5.9}$$

5.2.2 Channel Model

According to Chapter 2, the delay-d MIMO channel matrix between the i-th MS and the BS, $\mathbf{H}_i[d]$ can be written as

$$\mathbf{H}_i[d] = \sqrt{\frac{N_tN_r}{N_CN_P}}\sum_{c=1}^{N_C}\sum_{l=1}^{N_P}\alpha_{cl}^i g(dT_s - \tau_c - \tau_l)\mathbf{a}_{MS}(\theta_c^i - \gamma_l^i)\mathbf{a}_{BS}{}^H(\varphi_c^i - \phi_l^i),$$

where $\alpha_{cl}^i, c = 1, 2\cdots N_c, l = 1, 2\cdots N_P$ is the complex gain of the l-th ray in the c-th cluster of the i-th MS, which follows $\mathcal{CN}(0,1)$, while $\mathbf{a}_{\mathrm{MS}}(\theta_c^i)$ and $\mathbf{a}_{\mathrm{BS}}(\varphi_c^i)$ are the MS and BS array response vectors, respectively, θ_c^i and φ_c^i denote the AoA and AoD of the c-th cluster of the i-th MS, respectively. γ_l^i and ϕ_l^i denote relative AoA and AoD shift of the l-th ray of the i-th MS, respectively. The frequency channel response of MS i at subcarrier k, $\mathbf{H}_i[k]$, can be then expressed as

$$\mathbf{H}_i[k] = \sum_{d=0}^{D-1}\mathbf{H}_i[d]e^{-j\frac{2\pi k}{K}d}. \tag{5.10}$$

5.2.3 Problem Formulation

The main objective of the design is to maximize the overall spectral efficiency of the system under the total transmit power constraint and the constant modulus constraints on each entry of the analog beamformers, assuming perfect knowledge of

$\mathbf{H}_i[k]$. That is, we aim to find the optimal hybrid precoder at the BS and optimal hybrid combiner for each user by solving the following problems

$$\underset{\mathbf{F}_{RF},\mathbf{F}_{BB}[k],\left\{\mathbf{W}_{RF,i},\mathbf{W}_{BB,i}[k]\right\}_{i=1}^{N_u}}{\text{maximize}} \ \frac{1}{K}\sum_{k=1}^{K} R[k] \tag{5.11a}$$

$$\text{s.t.} \quad \text{Tr}\left(\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{F}_{BB}[k]^H\mathbf{F}_{RF}^H\right) \leq P \tag{5.11b}$$

$$\left|\mathbf{F}_{RF}(m,n)\right|^2 = 1, \forall m,n. \tag{5.11c}$$

$$\left|\mathbf{W}_{RF,i}(m,n)\right|^2 = 1, \forall m,n,i. \tag{5.11d}$$

To solve the above optimization problem, joint optimization on the RF and baseband precoders and combiners are required. However, in general, the optimization problem described in Equation (5.11a) is a non-convex problem due to the presence of the variables $\{\mathbf{F}_{RF}, \mathbf{F}_{BB}[k]\}$ and $\{\mathbf{W}_{RF}, \mathbf{W}_{BB,i}[k]\}$ in the interference plus noise term $(P_I[k] + P_N[k])$ and the product between the variables. In addition, as stated in [58], due to the non-convex constraints on $\mathbf{F}_{RF}$ and $\mathbf{W}_{RF,i}$, finding global optima for Equation (5.11a) is often found to be intractable. This paper uses the alternative strategy proposed in [60–62], in which the RF beamforming is first designed regardless of inter-user and intra-user interference, then the baseband beamforming is designed to cancel interference given the already designed RF beamforming.

5.3 Proposed Wideband Hybrid Beamforming Design

In this section, based on the wideband mmWave channel model and formulated problem in Section 5.2, we consider hybrid precoder and combiner design for wideband mmWave multi-user MIMO-OFDM systems. The RF beamforming aiming to maximize the mutual information or spectral efficiency is designed by a relaxation of the original maximization problem. Then the baseband beamforming is designed by SO-based SIC to cancel inter-user interference and ZF SIC to cancel intra-user interference. It should be noted that the receivers of multiple users need to have other users information and some sort of coordination among users is needed to do SIC. This is the price paid for SIC. The SIC receiver on the k-th subcarrier for desired MS i is shown in Figure 5.2. This strategy leads to a decoupled RF and baseband beamforming design. Following this strategy, the RF and baseband beamforming design

are given below.

5.3.1 RF beamforming design

As mentioned before, the RF beamforming is performed in the time domain and the same RF beamforming matrix is applied for the entire bandwidth. We propose to design RF beamforming by leveraging the sample covariance matrix of the frequency-domain channel.

RF precoder

According to Equation (5.2), the mutual information of MS i at subcarrier k achieved by Gaussian signaling over the mmWave channel is given by

$$\begin{aligned} I_i[k] &= \log_2 \left| \mathbf{I} + \frac{P_i}{\sigma^2} \mathbf{H}_i[k] \mathbf{F}_{RF} \mathbf{F}_{BB}[k] \mathbf{F}_{BB}^H[k] \mathbf{F}_{RF}^H \mathbf{H}_i^H[k] \right| \\ &\overset{(a)}{\leq} \log_2 \left| \mathbf{I} + \frac{P_i \gamma^2}{\sigma^2} \mathbf{H}_i[k] \mathbf{F}_{RF} \mathbf{F}_{RF}^H \mathbf{H}_i^H[k] \right| \\ &\overset{(b)}{=} \log_2 \left| \mathbf{I} + \frac{P_i \gamma^2}{\sigma^2} \mathbf{F}_{RF}^H \mathbf{H}_i^H[k] \mathbf{H}_i[k] \mathbf{F}_{RF} \right|. \end{aligned} \tag{5.12}$$

The simplification step (a) of Equation (5.12) holds due to $\mathbf{F}_{BB}[k]\mathbf{F}_{BB}^H[k] \approx \gamma^2 \mathbf{I}_{N_{RF}^t}$ [56] [96], and (a) is satisfied with equality if $N_{RF}^t = N_u N_S$. The simplification step (b) holds due to $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$. The mutual information achieved by all the MSs across all subcarriers is given by

$$\begin{aligned} I &= \frac{1}{K} \sum_{i=1}^{K_u} \sum_{k=1}^{K} I_i[k] \\ &= \frac{1}{K} \sum_{i=1}^{K_u} \sum_{k=1}^{K} \log_2 \left| \mathbf{I} + \frac{P_i \gamma^2}{\sigma^2} \mathbf{F}_{RF}^H \mathbf{H}_i^H[k] \mathbf{H}_i[k] \mathbf{F}_{RF} \right| \\ &\overset{(c)}{\leq} \log_2 \left| \mathbf{I} + \frac{\gamma^2}{\sigma^2} \mathbf{F}_{RF}^H \frac{1}{K} \sum_{i=1}^{K_u} \sum_{k=1}^{K} P_i \mathbf{H}_i^H[k] \mathbf{H}_i[k] \mathbf{F}_{RF} \right| \\ &= \log_2 \left| \mathbf{I} + \frac{\gamma^2}{\sigma^2} \mathbf{F}_{RF}^H \mathbf{C}_{ave} \mathbf{F}_{RF} \right|. \end{aligned} \tag{5.13}$$

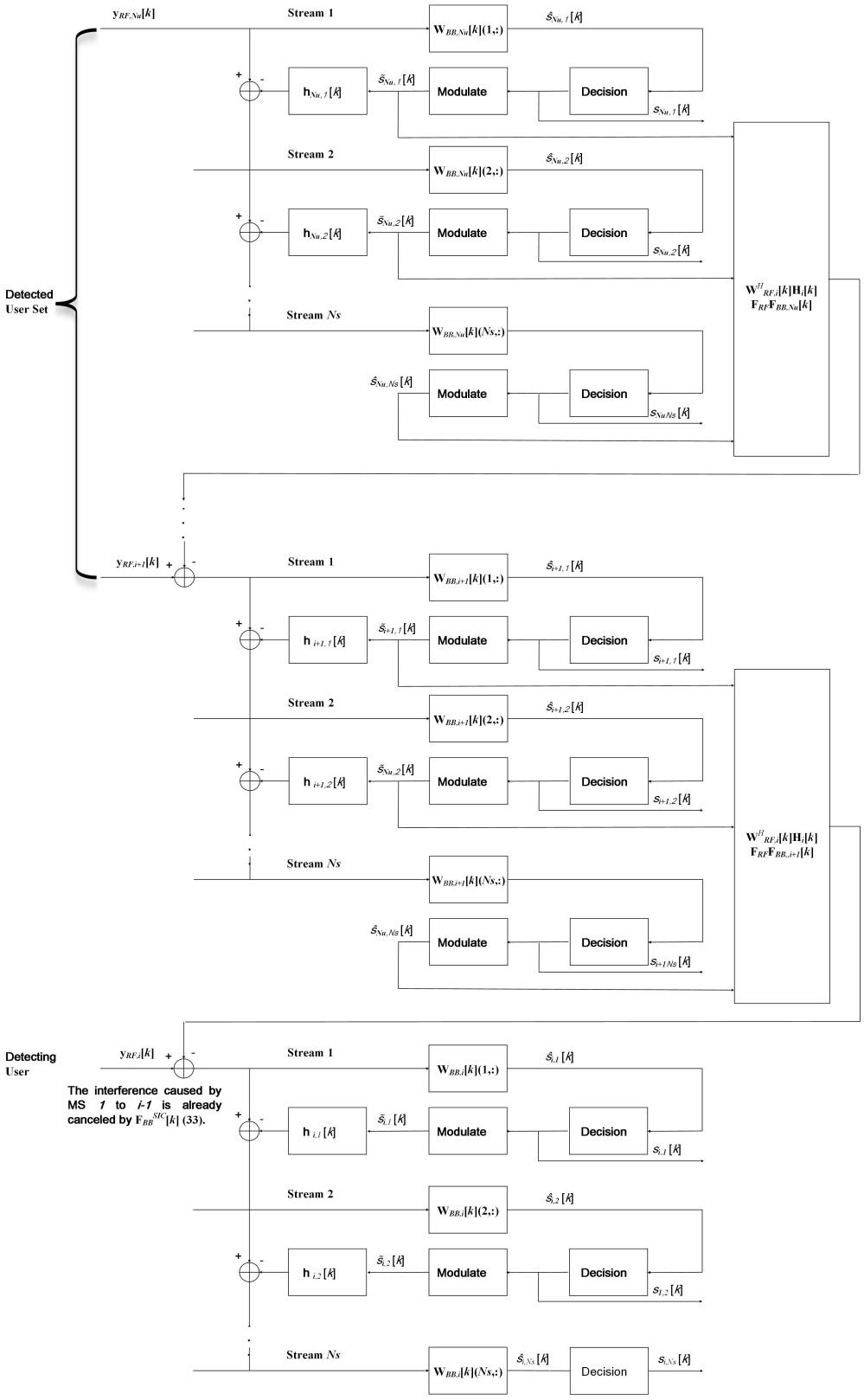

Figure 5.2: SIC receiver on the k-th subcarrier for desired MS i

The simplification step (c) of Equation (5.13) holds due to Jensen's inequality. $\mathbf{C}_{ave}$ stands for the average of the frequency-domain channel covariance matrix of all users, given by

$$\mathbf{C}_{ave} = \frac{1}{K}\sum_{i=1}^{K_u}\sum_{k=1}^{K} P_i \mathbf{H}_i^H[k]\mathbf{H}_i[k]. \tag{5.14}$$

When double the least number of RF chains are available at the transmitter, i.e., $N_{RF}^t = 2N_uN_s$, two RF chains with constant modulus constraints can be combined to act as one RF chain without constant modulus constraint. From Equation (5.13), the optimal unconstrained RF precoder is given by

$$\mathbf{F}_{RFun} = \mathbf{V}_{ave}(:, 1 : N_{RF}^t), \tag{5.15}$$

where $\mathbf{V}_{ave}(:, 1 : N_{RF}^t)$ consists of the first N_{RF}^t left singular vectors of $\mathbf{C}_{ave}$. When double the least number of RF chains are not available, i.e., $N_uN_s \leq N_{RF}^t < 2N_uN_s$, $\mathbf{V}_{ave}(:, 1 : N_{RF}^t)$ does not satisfy the constant modulus constraints. According to the prior work [58], we calculate the constrained RF precoder $\mathbf{F}_{RF}$ by minimizing the Frobenius norm between the optimal unconstrained and constrained RF precoder, which is given by

$$\begin{aligned} \mathbf{F}_{RF} &= \min_{\mathbf{F}_{RF}} \|\mathbf{F}_{RFun} - \mathbf{F}_{RF}\|_F^2 \\ s.t.\ &|\mathbf{F}_{RF}(i,j)| = 1/\sqrt{N_t}, \forall i, j. \end{aligned} \tag{5.16}$$

$\mathbf{F}_{RF}$ can be obtained by

$$\mathbf{F}_{RF} = \frac{1}{\sqrt{N_t}} e^{j\angle \mathbf{V}_{ave}(:,1:N_{RF}^t)}. \tag{5.17}$$

Proof: See Appendix B.

RF combiner

The received signal of MS i at subcarrier k after RF combining is given by

$$\begin{aligned}\mathbf{y}_{RF,i}[k] &= \mathbf{W}_{RF,i}^{H}\mathbf{r}_i[k] \\ &= \mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{S}[k] + \mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k].\end{aligned} \tag{5.18}$$

Without considering inter-user interference, the spectral efficiency achieved by MS i at subcarrier k is given by Equation (5.19), where (d) is based on $\mathbf{F}_{BB}[k]\mathbf{F}_{BB}^{H}[k] \approx \gamma^2\mathbf{I}_{N_{RF}^{t}}$ [56] [96].

$$\begin{aligned}R_i[k] &= \log_2\left|\mathbf{I} + \frac{P_i\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{F}_{BB}^{H}[k]\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}}{\sigma^2\mathbf{W}_{RF,i}^{H}\mathbf{W}_{RF,i}}\right| \\ &\overset{(d)}{\leq} \log_2\left|\mathbf{I} + \frac{P_i\gamma^2}{\sigma^2}\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}\right|.\end{aligned} \tag{5.19}$$

The average spectral efficiency achieved by MS i across all subcarriers R_i is given by Equation (5.20),

$$\begin{aligned}R_i &= \frac{1}{K}\sum_{k=1}^{K} R_i[k] = \frac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I} + \frac{P_i\gamma^2}{\sigma^2}\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}\right| \\ &\overset{(e)}{\leq} \log_2\left|\mathbf{I} + \frac{P_i\gamma^2}{\sigma^2}\mathbf{W}_{RF,i}^{H}\frac{1}{K}\sum_{k=1}^{K}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}\right| \\ &= \log_2\left|\mathbf{I} + \frac{P_i\gamma^2}{\sigma^2}\mathbf{W}_{RF,i}^{H}\mathbf{C}_{ave,i}\mathbf{W}_{RF,i}\right|.\end{aligned} \tag{5.20}$$

where (e) in Equation (5.20) is based on Jensen's inequality. $\mathbf{C}_{ave,i}$ represents the average of the frequency-domain equivalent channel covariance matrix of MS i, given by

$$\mathbf{C}_{ave,i} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]. \tag{5.21}$$

When double the least number of RF chains are available at the receiver, i.e., $N_{RF}^{r} = 2N_s$, the RF optimal unstrained combiner of MS i is given by

$$\mathbf{W}_{RFun,i} = \mathbf{V}_{ave,i}(:, 1 : N_{RF}^{r}), \tag{5.22}$$

where $\mathbf{V}_{ave,i}(:, 1 : N_{RF}^r)$ consists of the first N_{RF}^r left singular vectors of $\mathbf{C}_{ave,i}$. When double the least number of RF chains are not available at the receiver, i.e., $N_s \leq N_{RF}^r < 2N_s$, as mentioned in the RF precoder, the constrained RF combiner of MS i can be obtained by extracting the phase of the optimal unstrained RF combiner, given by

$$\mathbf{W}_{RF,i} = \frac{1}{\sqrt{N_r}} e^{j\angle \mathbf{V}_{ave,i}(:,1:N_{RF}^r)}. \tag{5.23}$$

The baseband equivalent channel of MS i at subcarrier k, $\tilde{\mathbf{H}}_i[k]$ with dimension $N_{RF}^r \times N_{RF}^t$ is given by

$$\tilde{\mathbf{H}}_i[k] = \mathbf{W}_{RF,i}^H \mathbf{H}_i[k] \mathbf{F}_{RF}, i = 1, 2, \cdots, N_u. \tag{5.24}$$

5.3.2 Baseband beamforming design

From the baseband combined signal Equation (5.3) and (5.4), both inter-user interference and intra-user interference need to be eliminated to demodulate the desired data stream. In this subsection, the baseband precoder and combiner are designed by employing the SIC-aided hybrid beamforming algorithm already proposed in [65] for single-carrier systems with frequency flat-fading channels to cancel inter-user and inner-user interference. Therefore, the baseband beamforming algorithms proposed in the following part will be specialized to the narrowband case if there is only one subcarrier. The rest of this subsection provides a brief description of the hybrid beamforming algorithm in [65] extended to OFDM-based wideband systems with frequency selective fading channels.

Baseband Precoder

To cancel inter-user interference, the baseband precoder of the i-th MS at subcarrier k, $\mathbf{F}_{BB,i}^{SIC}[k]$ needs to satisfy $\tilde{\mathbf{H}}_j[k]\mathbf{F}_{BB,i}^{SIC}[k] = \mathbf{0}, i+1 \leq j \leq N_u$. We define $\bar{\mathbf{H}}_i[k], (1 \leq k < K)$ as

$$\bar{\mathbf{H}}_i[k] = \left[\tilde{\mathbf{H}}_{i+1}^T[k], \tilde{\mathbf{H}}_{i+2}^T[k], \cdots, \tilde{\mathbf{H}}_{N_u}^T[k]\right]^T \tag{5.25}$$

Then the SVD of $\bar{\mathbf{H}}_i[k]$ is given by

$$\bar{\mathbf{H}}_i[k] = \mathbf{U}_i[k]\mathbf{\Sigma}_i[k]\mathbf{V}_i^H[k] = \mathbf{U}_i[k]\mathbf{\Sigma}_i[k][\mathbf{V}_i^1[k], \mathbf{V}_i^2[k]]^H, \tag{5.26}$$

where $\mathbf{V}_i^1[k]$ consists of the first $(N_u - i)N_{RF}^r$ columns of $\mathbf{V}_i[k]$, and $\mathbf{V}_i^2[k]$ holds the rest $N_{RF}^t - (N_u - i)\, N_{RF}^r$ columns of $\mathbf{V}_i[k]$, which are exactly the orthogonal bases of the null space of $\bar{\mathbf{H}}_i[k]$. Note that the null space of $\bar{\mathbf{H}}_i[k], 1 \leq i < N_u$ are not overlapped any more. Therefore, the SIC-based inter-user interference cancellation algorithm outperforms its linear counterparts. For the N_u-th MS, there is no interfering user according to our assumptions and we perform SVD of its baseband equivalent channel $\tilde{\mathbf{H}}_{N_u}[k]$, given by

$$\begin{aligned} \tilde{\mathbf{H}}_{N_u}[k] &= \mathbf{U}_{N_u}[k]\mathbf{\Sigma}_{N_u}[k]\mathbf{V}_{N_u}^H[k] \\ &= \mathbf{U}_{N_u}[k]\mathbf{\Sigma}_{N_u}[k]\left[\mathbf{V}_{N_u}^1[k], \mathbf{V}_{N_u}^2[k]\right]^H, \end{aligned} \tag{5.27}$$

where $\mathbf{V}_{N_u}^1[k]$ consists of the first N_s columns of $\mathbf{V}_{N_u}[k]$, and $\mathbf{V}_{N_u}^2[k]$ holds the rest $N_{RF}^t - N_s$ columns of $\mathbf{V}_{N_u}[k]$. Since the previous $N_u - 1$ MSs have been cancelled, the N_u-th MS can perform inter-user interference free multi-stream transmission at subcarrier k.

For the first $N_u - 1$ MSs, to achieve inter-user interference-free transmission, we force the baseband precoder $\mathbf{F}_{BB,i}^{SIC}[k]$ of MS i at subcarrier k lies in the null space of $\bar{\mathbf{H}}_i[k]$ by setting $\mathbf{F}_{BB,i}^{SIC}[k] = \mathbf{V}_i^2[k]\mathbf{G}_i[k], 1 \leq i < N_u$. For MS N_u, $\mathbf{F}_{BB,N_u}^{SIC}[k] = \mathbf{V}_{N_u}^1[k]\mathbf{G}_{N_u}[k]$. $\mathbf{G}_i[k], 1 \leq i \leq N_u$ is derived to maximize the spectral efficiency of the i-th MS at subcarrier k, given by

$$\mathbf{G}_i[k] = \mathbf{M}_i^{(N_s)}[k]\mathbf{\Gamma}_i^{1/2}[k]. \tag{5.28}$$

Proof: See Appendix C.

The baseband precoder $\mathbf{F}_{BB,i}^{SIC}[k]$ is summarized as

$$\mathbf{F}_{BB,i}^{SIC}[k] = \begin{cases} \mathbf{V}_i^2[k]\mathbf{G}_i[k] & 1 \leq i < N_u \\ \mathbf{V}_{N_u}^1[k]\mathbf{G}_{N_u}[k] & i = N_u. \end{cases} \tag{5.29}$$

Finally, the overall baseband precoder to enable inter-user interference-free transmis-

sion is given by

$$\mathbf{F}_{BB}^{SIC}[k] = \left[\mathbf{F}_{BB,1}^{SIC}[k], \mathbf{F}_{BB,2}^{SIC}[k], \cdots, \mathbf{F}_{BB,N_u}^{SIC}[k]\right]. \tag{5.30}$$

Obviously, the baseband precoder of the i-th MS at subcarrier k, $\mathbf{F}_{BB,i}^{SIC}[k]$ satisfies

$$\begin{aligned}\tilde{\mathbf{H}}_i[k]\mathbf{F}_{BB,j}^{SIC}[k] &= \mathbf{W}_{RF,i}^H[k]\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,j}^{SIC}[k] \\ &= \begin{cases} \mathbf{0} & 1 \le j < i \\ \mathbf{W}_{RF,i}^H[k]\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{V}_j^2[k]\mathbf{G}_j[k] & i \le j < N_u - 1 \\ \mathbf{W}_{RF,i}^H[k]\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{V}_{N_u}^1[k]\mathbf{G}_{N_u}[k] & j = N_u \end{cases}\end{aligned} \tag{5.31}$$

After the detected $(i+1)$-th to N_u-th MS has been cancelled, the RF received signal of the i-th MS at subcarrier k is given by

$$\begin{aligned}\bar{\mathbf{y}}_i[k] &= \mathbf{y}_{RF,i}[k] - \sum_{j=i+1}^{N_u} \mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,j}^{SIC}[k]\mathbf{s}_j[k] \\ &= \mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k]\mathbf{s}_i[k] + \mathbf{W}_{RF,i}^H\mathbf{n}_i[k] \\ &= \mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k](:,m)s_{i,m}[k] \\ &\quad + \textstyle\sum_{n=1,n\neq m}^{N_s} \mathbf{W}_{RF,i}^H\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k](:,n)s_{i,n}[k] \\ &\quad + \mathbf{W}_{RF,i}^H\mathbf{n}_i[k].\end{aligned} \tag{5.32}$$

The first equality of Equation (5.32) holds due to Equation (5.31). The three successive terms of the third equality of Equation (5.32) on different lines represent the m-th desired data stream for MS i at subbcarrier k, the intra-user interference and the effective noise, respectively.

Baseband Combiner

Without loss of generality, we want to detect the m-th stream of the i-th MS at subcarrier k. Define the equivalent channel of the i-th MS at subcarrier k as $\mathbf{H}_{Equ,i}[k]$, given by

$$\mathbf{H}_{Equ,i}[k] = \tilde{\mathbf{H}}_i[k]\mathbf{F}_{BB,i}^{SIC}[k] = \left[\mathbf{h}_{i,1}[k], \cdots, \mathbf{h}_{i,N_s}[k]\right]. \tag{5.33}$$

The ZF baseband combiner to detect the m-th stream of MS i at subcarrier k is $\mathbf{W}_{BB,i}^{H}[k](m,:) = \mathbf{G}_k(m,:)$, where $\mathbf{G}_k$ is the ZF baseband combiner, given by

$$\mathbf{G}_k = \left(\mathbf{H}_{Equ,i}^{H}[k]\mathbf{H}_{Equ,i}[k]\right)^{-1}\mathbf{H}_{Equ,i}[k]. \tag{5.34}$$

Therefore, the detected m-th stream of the MS i at subcarrier k is given by

$$\begin{aligned}\hat{s}_{i,m}[k] &= \mathbf{W}_{BB,i}^{H}[k](m,:)\bar{\mathbf{y}}_i[k] \\ &= \mathbf{W}_{BB,i}^{H}[k](m,:)\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k](:,m)s_{i,m}[k] \\ &\quad + \mathbf{W}_{BB,i}^{H}[k](m,:)\mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k].\end{aligned} \tag{5.35}$$

Then $\hat{s}_{i,m}[k]$ is demodulated and re-modulated to cancel the noise accumulation. After that the re-modulated symbol $\tilde{s}_{i,m}[k]$ will be subtracted from the received signal, which is given by

$$\begin{aligned}\bar{\mathbf{y}}_{i,m}[k] &= \bar{\mathbf{y}}_i[k] - \mathbf{h}_{i,m}[k]\tilde{s}_{i,m}[k] \\ &= \sum_{q=1,q\neq m}^{N_s}\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k](:,q)s_{i,q}[k] \\ &\quad + \mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k].\end{aligned} \tag{5.36}$$

Then we will detect the n-th ($n \neq m$) data stream of the i-th MS at subcarrier k, $s_{i,n}[k]$. Let $\bar{\mathbf{H}}_{i,m}^{SIC}[k]$ denote zeroing column m of $\mathbf{H}_{Equ,i}[k]$, which is given by

$$\bar{\mathbf{H}}_{i,m}^{SIC}[k] = \left[\mathbf{h}_{i,1}[k], \cdots, \mathbf{h}_{i,m-1}[k], \mathbf{0}, \mathbf{h}_{i,m+1}[k], \cdots, \mathbf{h}_{i,N_s}[k]\right]. \tag{5.37}$$

To detect the n-th data stream of the i-th MS at subcarrier k, $s_{i,n}[k]$, the baseband combiner $\mathbf{W}_{BB,i}^{H}[k](n,:) = \mathbf{Q}_k(n,:)$, where $\mathbf{Q}_k$ is given by

$$\mathbf{Q}_k = \left(\left(\bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^{H}\bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^{-1}\left(\bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^{H}. \tag{5.38}$$

The detected n-th data stream of MS i at subcarrier k, $\hat{s}_{i,n}[k]$ is given by

$$\begin{aligned}\hat{s}_{i,n}[k] &= \mathbf{W}_{BB,i}^{H}[k](n,:)\bar{\mathbf{y}}_{i,m}[k] \\ &= \mathbf{W}_{BB,i}^{H}[k](n,:)\sum_{q=1,q\neq m}^{N_s}\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}(:,q)s_{i,q}[k] \\ &\quad + \mathbf{W}_{BB,i}^{H}[k](n,:)\mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k] \\ &\overset{(f)}{=} \mathbf{W}_{BB,i}^{H}[k](n,:)\mathbf{W}_{RF,k}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}(:,n)[k]s_{i,n}[k] \\ &\quad + \mathbf{W}_{BB,i}^{H}[k](n,:)\mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k].\end{aligned} \tag{5.39}$$

The simplification of step (f) of Equation (5.39) holds due to Equation (5.38). Then the detected symbol $s_{i,m}[k]$ and $s_{i,n}[k]$ will be subtracted from the received signal, and the above steps are then performed for the rest streams by operating in turn on the progression of the already-detected symbol removed. Since the preceding treatment eventually cancels both inter-user and intra-user interference, the sum spectral efficiency of all MSs across all subcarriers can be simplified as Equation (5.40). In summary, we provide the proposed hybrid SIC algorithm as Algorithm 9.

$$R=\sum_{k=1}^{K}\sum_{i=1}^{N_u}\sum_{n=1}^{N_s}\frac{P_i}{\sigma^2 K N_s}\frac{\mathbf{W}_{BB,i}^{H}[k](n,:)\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k](:,n)\mathbf{F}_{BB,i}{}^{H}[k](:,n)\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:)}{\mathbf{W}_{BB,i}^{H}[k](n,:)\mathbf{W}_{RF,i}^{H}\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](n,:)}. \tag{5.40}$$

Detection User and stream Order

As discussed in [65], the order in which user are detected can be determined by their large scale fading, and the order in which streams of the i-th MS $\mathbf{s}_i$ are detected according to the best post-detection SINR of the m-th stream of the i-th MS $SINR_{i,m}$ at each stream detection, as given by Equation (5.41). Note that $P_I[k]$ in Equation (5.42) is different from Equation 5.8 since the inter-user interference has been cancelled in Equation (5.35).

$$SINR_{i,m}[k]=\frac{P_i\mathbf{W}_{BB,i}^{H}[k](m,:)\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k](m,:)\big(\mathbf{F}_{BB,i}^{SIC}[k](m,:)\big)^{H}\mathbf{F}_{RF}^{H}\mathbf{H}_i^{H}[k]\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](m,:)}{\sigma^2\mathbf{W}_{BB,i}^{H}[k](m,:)\mathbf{W}_{RF,i}^{H}\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](m,:)+P_I} \tag{5.41}$$

$$P_I[k]=\mathbf{W}_{BB,i}^{H}[k](m,:)\mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\sum_{q=1,q\neq m}^{N_s}P_i\mathbf{F}_{BB,i}^{SIC}[k](q,:)\big(\mathbf{F}_{BB,i}^{SIC}[k](q,:)\big)^{H}\mathbf{F}_{RF}^{H}\mathbf{H}_i[k]^{H}\mathbf{W}_{RF,i}\mathbf{W}_{BB,i}[k](m,:) \tag{5.42}$$

Algorithm 9 Hybrid SIC algorithm

Input: $\mathbf{H}_i[k], i = 1, 2, \cdots N_u; k = 1, 2, \cdots K.$
Analog Stage
1. **for** $i \leq N_u$ **do**
2. **for** $k \leq K$ **do**
3. Calculate $\mathbf{C}_{ave}$ according to Equation (5.14)
4. **end for**
5. **end for**
6. Calculate $\mathbf{F}_{RF}$ according to Equation (5.15) or (5.17)
7. **for** $i \leq N_u$ **do**
8. **for** $k \leq K$ **do**
9. Calculate $\mathbf{C}_{ave,i}$ according to Equation (5.21)
10. **end for**
11. Calculate $\mathbf{W}_{RF,i}$ according to Equation (5.22) or Equation (5.23)
12. **end for**
13. $i = 1, 2, \cdots N_u; k = 1, 2, \cdots K,$
Calculate $\tilde{\mathbf{H}}_i[k]$ according to Equation (5.24)
End Analog Stage
Digital Stage
14. **for** $k \leq K$ **do**
15. **for** $i \leq N_u$ **do**
16. **if** $i < N_u$ **do**
17. $\bar{\mathbf{H}}_i[k] = \left[\tilde{\mathbf{H}}_{i+1}^T[k], \tilde{\mathbf{H}}_{i+2}^T[k], \cdots, \tilde{\mathbf{H}}_{N_u}^T[k]\right]^T$
18. $\bar{\mathbf{H}}_i[k] = \mathbf{U}_i[k]\mathbf{\Sigma}_i[k]\left[\mathbf{V}_i^1[k], \mathbf{V}_i^2[k]\right]^H$
19. $\mathbf{F}_{BB,i}^{SIC}[k] = \mathbf{V}_i^2[k]\mathbf{G}_i[k]$
20. **else**
21. $\tilde{\mathbf{H}}_{N_u}[k] = \mathbf{U}_{N_u}[k]\mathbf{\Sigma}_{N_u}[k]\left[\mathbf{V}_{N_u}^1[k], \mathbf{V}_{N_u}^2[k]\right]^H$
22. $\mathbf{F}_{BB,N_u}^{SIC}[k] = \mathbf{V}_{N_u}^2\mathbf{G}_{N_u}[k]$
23. **end if**
24. **end for**
25. **Obtain** $\mathbf{F}_{BB}^{SIC}[k] = \left[\mathbf{F}_{BB,1}^{SIC}[k], \cdots, \mathbf{F}_{BB,N_u}^{SIC}[k]\right]$
26. **end for**
27. **for** $k \leq K$ **do**
28. **for** $i \leq N_u$ **do**
29. Calculate $\mathbf{H}_{Equ,i}[k]$ according to Equation (5.33)
30. **for** $m \leq N_s$ **do**
31. Calculate $\bar{\mathbf{H}}_{i,m}^{SIC}[k]$ according to Equation (5.37)
32. $\mathbf{Q}_k = \left(\left(\bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^H \bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^{-1} \left(\bar{\mathbf{H}}_{i,m}^{SIC}[k]\right)^H$
33. $\mathbf{W}_{BB,i}[k](m,:) = \mathbf{Q}_k\,(m,:)$
34. **end for**
35. **end for**
36. **end for**
End Digital Stage
Output $\mathbf{F}_{RF}, \mathbf{F}_{BB}^{SIC}[k], \mathbf{W}_{RF,i}, \mathbf{W}_{BB,i}[k],$
$i = 1, 2, \cdots, N_u; k = 1, 2, \cdots, K$

5.4 Channel Estimation Errors

In Section III, we assume that perfect channel state information is available. In this section, we analyze the performance of the proposed hybrid beamforming with

imperfect CSI. According to Chapter 3, the estimated channel matrix of MS i at subcarrier k can be modeled as [17]

$$\hat{\mathbf{H}}_i[k] = \xi \mathbf{H}_i[k] + \sqrt{1-\xi^2}\mathbf{E}_i[k], \tag{5.43}$$

where $0 \leq \xi \leq 1$ represents the reliability of the estimate and $\mathbf{E}_i[k]$ is the estimation error matrix that is uncorrelated with $\mathbf{H}_i[k]$, the entries of $\mathbf{E}_i[k]$ are i.i.d and follow a complex Gaussian distribution of $\mathcal{CN}(0, \mathbf{I})$. For the RF precoding, substituting Equation (5.43) into (5.13), the mutual information can be expressed as Equation (5.44), where the third equality holds due to the fact that $\mathbf{H}_i[k]$ and $\mathbf{E}_i[k]$ are uncorrelated, and the last line is obtained based on the fact that each element of $\mathbf{E}_i[k]$ follows a complex Gaussian distribution of $\mathcal{CN}(0,1)$ and K is large.

$$\begin{aligned}
I &= \tfrac{1}{K}\sum_{i=1}^{N_u}\sum_{k=1}^{K}\log_2\left|\mathbf{I} + \tfrac{P_i}{\sigma^2}\mathbf{F}_{RF}^H\left(\xi\mathbf{H}_i[k] + \sqrt{1-\xi^2}\mathbf{E}_i[k]\right)^H\left(\xi\mathbf{H}_i[k] + \sqrt{1-\xi^2}\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\right| \\
&= \tfrac{1}{K}\sum_{i=1}^{N_u}\sum_{k=1}^{K}\log_2\left|\mathbf{I} + \tfrac{P_i}{\sigma^2}\mathbf{F}_{RF}^H\left(\xi^2\mathbf{H}_i^H[k]\mathbf{H}_i[k] + 2\mathrm{Re}\left(\xi\sqrt{1-\xi^2}\mathbf{E}_i^H[k]\mathbf{H}_i[k]\right) + \left(1-\xi^2\right)\mathbf{E}_i^H[k]\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\right| \\
&\leq \log_2\left|\mathbf{I} + \tfrac{P_i}{\sigma^2}\mathbf{F}_{RF}^H\tfrac{1}{K}\sum_{i=1}^{N_u}\sum_{k=1}^{K}\left(\xi^2\mathbf{H}_i^H[k]\mathbf{H}_i[k] + 2\mathrm{Re}\left(\xi\sqrt{1-\xi^2}\mathbf{E}_i^H[k]\mathbf{H}_i[k]\right) + \left(1-\xi^2\right)\mathbf{E}_i^H[k]\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\right| \\
&= \log_2\left|\mathbf{I} + \tfrac{P_i}{\sigma^2}\mathbf{F}_{RF}^H\tfrac{1}{K}\sum_{i=1}^{N_u}\sum_{k=1}^{K}\left(\xi^2\mathbf{H}_i^H[k]\mathbf{H}_i[k] + \left(1-\xi^2\right)\mathbf{E}_i^H[k]\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\right| \\
&\approx \log_2\left|\mathbf{I} + \tfrac{P_i}{\sigma^2}\mathbf{F}_{RF}^H\left(\tfrac{\xi^2}{K}\sum_{i=1}^{N_u}\sum_{k=1}^{K}\mathbf{H}_i^H[k]\mathbf{H}_i[k]\right)\mathbf{F}_{RF}\right|
\end{aligned} \tag{5.44}$$

For RF combining, substituting Equation (5.43) into (5.20), as in RF precoding, the average spectral efficiency of MS i can be further expressed as Equation (5.45). From Equation (5.44) and (5.45), we can observe that $\frac{\xi^2}{K}\sum_{i=1}^{K_u}\sum_{k=1}^{K}\mathbf{H}_i^H[k]\mathbf{H}_i[k]$ and $\mathbf{C}_{ave}$ have the same left singular vectors. For the same reason, $\frac{\xi^2}{K}\sum_{k=1}^{K}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]$ and $\mathbf{C}_{ave,i}$ have the same left singular vectors. Therefore, when the number of subcarriers and ξ is large, the proposed RF beamforming is virtually unaffected by channel

estimation error.

$$
\begin{aligned}
R_i &= \tfrac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}+\tfrac{P_i}{\sigma^2}\mathbf{W}_{RF,i}^H\left(\xi\mathbf{H}_i[k]+\sqrt{1-\xi^2}\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\mathbf{F}_{RF}^H\left(\xi\mathbf{H}_i[k]+\sqrt{1-\xi^2}\mathbf{E}_i[k]\right)^H\mathbf{W}_{RF,i}\right| \\
&= \tfrac{1}{K}\sum_{k=1}^{K}\log_2\left|\mathbf{I}+\tfrac{P_i}{\sigma^2}\mathbf{W}_{RF,i}^H\left(\begin{array}{l}\xi^2\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]+2\mathrm{Re}\left(\xi\sqrt{1-\xi^2}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{E}_i^H[k]\right)\\ +\left(1-\xi^2\right)\mathbf{E}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{E}_i^H[k]\end{array}\right)\mathbf{W}_{RF,i}\right| \\
&\le \log_2\left|\mathbf{I}+\tfrac{P_i}{\sigma^2}\mathbf{W}_{RF,i}^H\tfrac{1}{K}\sum_{k=1}^{K}\left(\begin{array}{l}\xi^2\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]+2\mathrm{Re}\left(\xi\sqrt{1-\xi^2}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{E}_i^H[k]\right)\\ +\left(1-\xi^2\right)\mathbf{E}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{E}_i^H[k]\end{array}\right)\mathbf{W}_{RF,i}\right| \\
&= \log_2\left|\mathbf{I}+\tfrac{P_i}{\sigma^2}\mathbf{W}_{RF,i}^H\tfrac{1}{K}\sum_{k=1}^{K}\left(\xi^2\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]+\left(1-\xi^2\right)\mathbf{E}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{E}_i^H[k]\right)\mathbf{W}_{RF,i}\right| \\
&\approx \log_2\left|\mathbf{I}+\tfrac{P_i}{\sigma^2}\mathbf{W}_{RF,i}^H\left(\tfrac{\xi^2}{K}\sum_{k=1}^{K}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{RF}^H\mathbf{H}_i^H[k]\right)\mathbf{W}_{RF,i}\right|
\end{aligned}
\tag{5.45}
$$

For the baseband precoding, substituting Equation (5.43) into (5.24), we can obtain $\hat{\bar{\mathbf{H}}}_i[k]$, which is given by Equation (5.46). $\bar{\mathbf{E}}_i[k]$ in Equation (5.46) is given by Equation (5.47). Defining the SVD of $\hat{\bar{\mathbf{H}}}_i[k]=\sum_{p=1}^{r_{i,k}}\lambda_p\mathbf{u}_p\mathbf{v}_p$, where $\mathbf{u}_p$ and $\mathbf{v}_p$ are the left and right singular vectors corresponding to the p-th singular value λ_p, $r_{i,k}\le\min\left(\left(N_u-i\right)N_{RF}^r,N_{RF}^t\right)$ is the rank of $\hat{\bar{\mathbf{H}}}_i[k]$. It is obvious that $\xi\bar{\mathbf{H}}_i[k]$ and $\bar{\mathbf{H}}_i[k]$ have the same null space, since they have the same right singular vectors corresponding to zero singular values. Moreover, if ξ is large enough, there exists a l_k that satisfies $\sum_{p=1}^{l_k}\lambda_p\mathbf{u}_p\mathbf{v}_p\approx\xi\bar{\mathbf{H}}_i[k]$. Therefore, we conclude that if ξ is large enough, $\hat{\bar{\mathbf{H}}}_i[k]$ and $\bar{\mathbf{H}}_i[k]$ have the same null space. That is to say, our proposed null space based baseband precoding design can reduce the influence of channel estimation errors.

$$
\begin{aligned}
\hat{\bar{\mathbf{H}}}_i[k]=&\left[\begin{array}{c}\mathbf{W}_{RF,i+1}^H\left(\xi\mathbf{H}_{i+1}+\sqrt{1-\xi^2}\mathbf{E}_{i+1}[k]\right)[k]\mathbf{F}_{RF}\\ \mathbf{W}_{RF,i+2}^H\left(\xi\mathbf{H}_{i+2}+\sqrt{1-\xi^2}\mathbf{E}_{i+2}[k]\right)[k]\mathbf{F}_{RF}\\ \vdots\\ \mathbf{W}_{RF,N_u}^H\left(\xi\mathbf{H}_{N_u}+\sqrt{1-\xi^2}\mathbf{E}_{N_u}[k]\right)[k]\mathbf{F}_{RF}\end{array}\right] \\
=&\left[\begin{array}{c}\mathbf{W}_{RF,i+1}^H\xi\mathbf{H}_{i+1}\mathbf{F}_{RF}\\ \mathbf{W}_{RF,i+2}^H\xi\mathbf{H}_{i+2}\mathbf{F}_{RF}\\ \vdots\\ \mathbf{W}_{RF,N_u}^H\xi\mathbf{H}_{N_u}\mathbf{F}_{RF}\end{array}\right]+\left[\begin{array}{c}\mathbf{W}_{RF,i+1}^H\sqrt{1-\xi^2}\mathbf{E}_{i+1}[k]\mathbf{F}_{RF}\\ \mathbf{W}_{RF,i+2}^H\sqrt{1-\xi^2}\mathbf{E}_{i+2}[k]\mathbf{F}_{RF}\\ \vdots\\ \mathbf{W}_{RF,N_u}^H\sqrt{1-\xi^2}\mathbf{E}_{N_u}[k]\mathbf{F}_{RF}\end{array}\right], \\
=&\ \xi\bar{\mathbf{H}}_i[k]+\sqrt{1-\xi^2}\bar{\mathbf{E}}_i[k]
\end{aligned}
\tag{5.46}
$$

$$
\bar{\mathbf{E}}_i[k]=\left[\left(\mathbf{W}_{RF,i+1}^H\mathbf{E}_{i+1}[k]\mathbf{F}_{RF}\right)^T,\left(\mathbf{W}_{RF,i+2}^H\mathbf{E}_{i+2}[k]\mathbf{F}_{RF}\right)^T,\cdots\left(\mathbf{W}_{RF,N_u}^H\mathbf{E}_{N_u}[k]\mathbf{F}_{RF}\right)^T\right]^T.
\tag{5.47}
$$

For the baseband combining, $\mathbf{G}_k$ is the pseudo-inverse of $\mathbf{H}_{Equ,i}[k]$, substituting Equation (5.43) into (5.33), we can obtain $\hat{\mathbf{H}}_{Equ,i}[k]$, given by

$$\begin{aligned}\hat{\mathbf{H}}_{Equ,i}[k] &= \mathbf{W}_{RF,i}^{H}\left(\xi\mathbf{H}_i[k]+\sqrt{1-\xi^2}\mathbf{E}_i[k]\right)\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k] \\ &= \xi\mathbf{H}_{Equ,i}[k]+\sqrt{1-\xi^2}\mathbf{W}_{RF,i}^{H}\mathbf{E}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k] \\ &= \xi\mathbf{H}_{Equ,i}[k]+\sqrt{1-\xi^2}\mathbf{Q}_{Equ,i}[k],\end{aligned} \tag{5.48}$$

where $\mathbf{Q}_{Equ,i}[k]=\mathbf{W}_{RF,i}^{H}\mathbf{E}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}[k]$. Assume ξ is very large, $\sqrt{1-\xi^2}$ is very small, then the pseudo-inverse of the equivalent channel matrix with estimation error can be approximated by the linear part of the Taylor expansion [127], which can be expressed as

$$\begin{aligned}\hat{\mathbf{H}}_{Equ,i}^{\dagger}[k] &= \left(\xi\mathbf{H}_{Equ,i}[k]+\sqrt{1-\xi^2}\mathbf{Q}_{Equ,i}[k]\right)^{\dagger} \\ &\approx (\xi\mathbf{H}_{Equ,i}[k])^{\dagger}\left(I-\sqrt{1-\xi^2}\mathbf{Q}_{Equ,i}[k](\xi\mathbf{H}_{Equ,i}[k])^{\dagger}\right) \\ &= \tfrac{1}{\xi}\mathbf{H}_{Equ,i}^{\dagger}[k]-\tfrac{\sqrt{1-\xi^2}}{\xi^2}\mathbf{H}_{Equ,i}^{\dagger}[k]\mathbf{Q}_{Equ,i}[k]\mathbf{H}_{Equ,i}^{\dagger}[k].\end{aligned} \tag{5.49}$$

From Equation (5.49), we can observe that for baseband combining, the effect of channel estimation errors is negligible when ξ is large enough.

5.5 Computational Complexity Analysis

In this section, we analyze the complexity of implementing the proposed hybrid SIC algorithm. For comparison, we also present the complexity in implementing SU fully digital SVD, the fully digital BD, extended RF phase extraction hybrid BD (Extended PEHBD) [62] and hybrid SLNR (HSLNR) [124]. To focus on the main ideas, we consider the simplified case where each MS is equipped with the same number of antennas N_r and RF chains N_{RF}^{r}. Furthermore, each MS is assumed to support the same number of data streams N_s.

For the SU fully digital SVD beamforming, the total complexity is dominated by K SVDs of the overall channel matrix with dimension $N_uN_r\times N_t$, yielding the complexity order of $\mathcal{O}(KN_u^2N_r^2N_t)$. For the fully digital BD beamforming, the total complexity is dominated by KN_u SVDs of the overall channel matrix with dimension $(N_u-1)N_r\times N_t$, yielding the complexity order of $\mathcal{O}(KN_u(N_u-1)^2N_r^2N_t)$. For the extended PEHBD, the total complexity is dominated by RF beamforming, and therefore can be shown to be the complexity order of $\mathcal{O}(N_t^3)$. For the HSLNR algorithm, the total

complexity is dominated by N_u times of SVDs of the horizontal unfolding of the frequency domain channel, yielding the complexity order of $\mathcal{O}(N_u K^3 N_t^3)$. For the proposed hybrid beamforming algorithm, the total complexity is dominated by SVD of $\mathbf{C}_{ave}$, yielding the complexity order of $\mathcal{O}(N_t^3)$.

For the baseband beamforming, the extra computation of the proposed algorithm compared to the existing linear algorithms when performing intra-user interference cancellation is dominated by KN_uN_s times of matrix inverse of $\hat{\mathbf{H}}_k$, and therefore can be shown to be of complexity $\mathcal{O}(KN_u^3(N_{RF}^r)^2N_{RF}^t) + \mathcal{O}(KN_uN_{RF}^rN_s^3) \approx \mathcal{O}(KN_u^3(N_{RF}^r)^2N_{RF}^t)$. Furthermore, the required extra complexity of the proposed algorithm compared to the existing linear algorithms when performing inter-user interference cancellation is dominated by $K(N_u-1)$ times of SVDs, yielding the complexity order of $\mathcal{O}(\frac{1}{3}KN_u^3(N_{RF}^r)^2N_{RF}^t)$, KN_u times of SVDs with the complexity order of $\mathcal{O}(KN_u^3(N_{RF}^r)^2N_{RF}^t)$ and KN_u times of matrix inverse with the complexity order of $\mathcal{O}(KN_u(N_{RF}^r)^2N_{RF}^t)$ when maximizing the spectral efficiency of each MS. Therefore the complexity order of the proposed algorithm is simplified to $\mathcal{O}(\frac{4}{3}KN_u^3(N_{RF}^r)^2N_{RF}^t)$. Therefore, the baseband beamforming complexity of the proposed hybrid SIC algorithm is higher than the extended PFHBD and HSLNR algorithms. In summary, we list the above complexity evaluation results in Table 5.1.

Table 5.1: Comparison of computational complexity

Algorithms	$\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$	$\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$	**Total Complexity**
Fullly Digital SVD	Not Involved	$\mathcal{O}(KN_u^2N_r^2N_t)$	$\mathcal{O}(KN_u^2N_r^2N_t)$
Fullly Digital BD	Not Involved	$\mathcal{O}(KN_u(N_u-1)^2N_r^2N_t)$	$\mathcal{O}(KN_u(N_u-1)^2N_r^2N_t)$
Extended PFHBD	$\mathcal{O}(N_t^3)$	$\mathcal{O}(KN_u(N_u-1)^2(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(N_t^3)$
HSLNR	$\mathcal{O}(KN_uN_r^2N_t)$	$\mathcal{O}(KN_u(N_u-1)^2(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(KN_uN_r^2N_t)$
Proposed Hybrid SIC	$\mathcal{O}(N_t^3)$	$\mathcal{O}(\frac{4}{3}KN_u^3(N_{RF}^r)^2N_{RF}^t)$	$\mathcal{O}(N_t^3)$

5.6 Simulation Results

In this section, we present simulation results to illustrate the sum SE performance of the proposed hybrid SIC algorithm. In the simulations, the least number of RF chains is considered at transceivers. Further, unless otherwise mentioned, both BS and MS are equipped with ULA with the number of antennas $N_t = 256$ and $N_r = 16$, respectively. The system carrier frequency is set to $f_c = 28$ GHz, and the bandwidth is set to $B = 4$ GHz. The total number of subcarriers is chosen as $K = 512$. The CP length L_{cp} is $K/4 = 128$ and the maximum delay is less than $L_{cp}T_s$. The channels

from BS to each MS with the same number of scattering clusters $N_c = 8$ and paths $N_p = 10$ are generated according to Equation (5.10) in which the center AoAs and AoDs of the clusters are assumed to be uniformly distributed in $[0, 2\pi]$ and the relative AoAs and AoDs are assumed to follow Laplacian distribution [85]. The path delay is assumed to be uniformly distributed in $[0, L_{cp}T_s)$. The antenna element spacing is set to $d_e = \lambda/2$. In simulations, the sum SE is plotted versus the SNR per subcarrier (SNR $= P/\sigma_n^2$) or the number of antenna elements or the number of users averaged over 5000 random channel realizations.

5.6.1 Sum Spectral Efficiency Evaluation

First, we numerically illustrate the sum SE performance of the proposed Hybrid SIC algorithm with different configurations in Figure 5.3 to 5.5. We compare the proposed hybrid SIC algorithm with the extended PEHBD [62], HSLNR [124] and fully digital BD scheme [101]. The benchmark is provided by the sum of the maximum achievable spectral efficiency of the SU fully digital SVD beamforming [58] with the same system configuration. Furthermore, we also provide an upper bound spectral efficiency of the proposed algorithm, given by Equation (5.20).

Three cases, $N_u = 8$, $N_u = 4$ and $N_u = 2$ are considered in Figure 5.3 to 5.5, respectively, and we suppose $N_{RF}^r = N_s = 4$, $N_{RF}^t = N_u N_s$. We can observe from Figure 5.3 to 5.5 that the proposed hybrid SIC algorithm achieves slightly lower sum spectral efficiency compared to the upper bound given by Equation (5.20), which confirms our analysis in Section 5.3. Furthermore, the spectral efficiency of the proposed hybrid SIC algorithm performs close to the SU fully digital SVD beamforming, and outperforms the fully digital BD algorithm in the case of $N_u = 8$ and $N_u = 4$. Moreover, with the more powerful nonlinear beamforming, the proposed hybrid SIC algorithm exhibits substantially higher sum spectral efficiency compared to the traditional linear hybrid beamforming algorithms.

Next, we evaluate the sum spectral efficiency when varying the number of BS antennas and MS antennas with SNR = 10 dB. In Figure 5.6, the same setup of Figure 5.4 is adopted except the number of ULA element at BS changes from $N_t = 50$ to $N_t = 400$. In Figure 5.7, the same setup of Figure 5.4 is adopted except the number of ULA element at MS changes from $N_r = 10$ to $N_r = 100$. The figures show a considerable performance gain by the proposed hybrid SIC algorithm over the extended PEHBD and HSLNR algorithm. The reason is that the inter-user

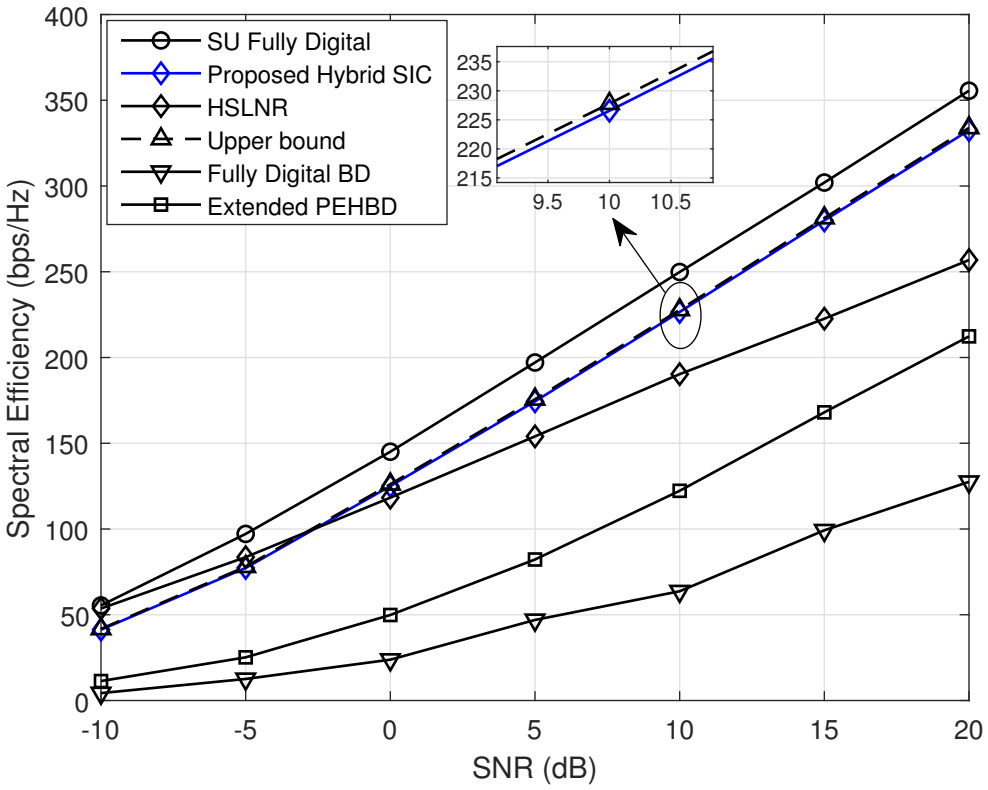

Figure 5.3: Sum SE performance comparison of different algorithms with $N_u = 8$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 32$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

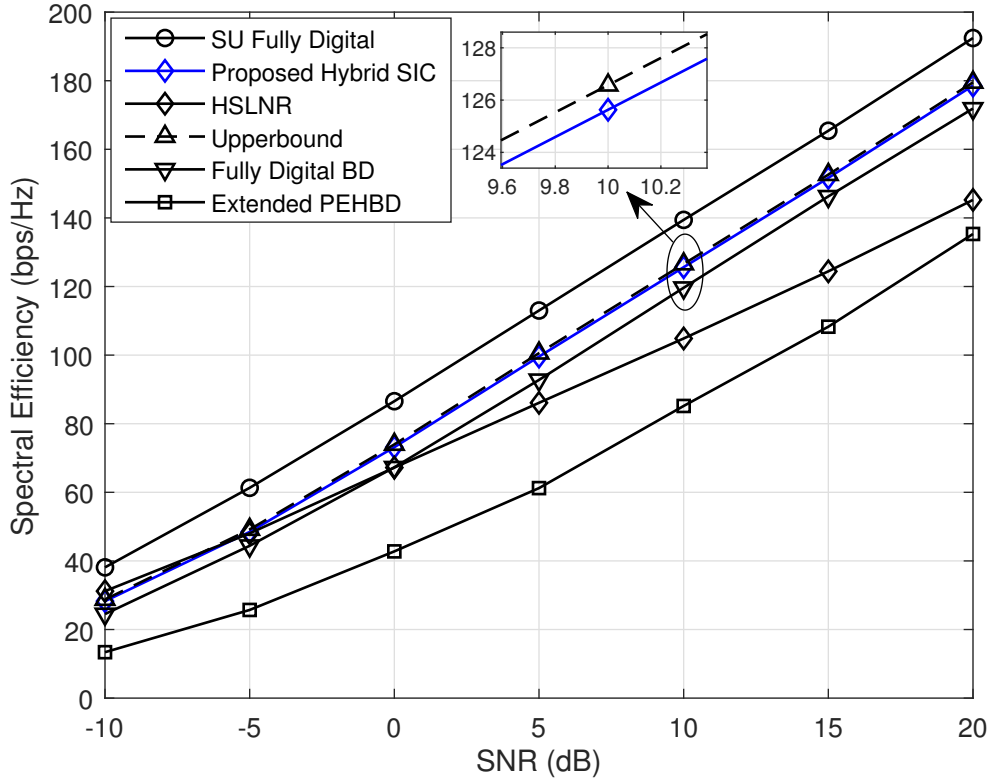

Figure 5.4: Sum SE performance comparison of different algorithms with $N_u = 4$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

interference and intra-user interference are both cancelled by SIC in the proposed hybrid SIC algorithm, which boosts the sum spectral efficiency performance.

Then we evaluate the performance of the proposed hybrid SIC algorithm with a different number of MS in Figure 5.8. In this figure, the same system setup is adopted in Figure 5.3, but with N_u increasing from 2 to 16, and $N_{RF}^t = N_u N_{RF}^r$. In this scenario, the sum spectral efficiency of the proposed hybrid SIC and HSLNR algorithm increase as the number of MSs N_u grows. A considerable performance improvement is observed by the proposed algorithm over the one in the literature. The reason is that inter-user interference is cancelled by SIC in the proposed algorithms, which boosts the sum spectral efficiency performance.

In Figure 5.9, we further illustrates the sum spectral efficiency improvement achieved by the proposed hybrid SIC algorithm with optimal data stream order (referred to as hybrid OSIC) in different cases. Three cases, $N_u = 8$, $N_u = 4$ and $N_u = 2$ are considered. Each MS is equipped with $N_{RF}^r = 4$ RF chains and supports $N_s = 4$

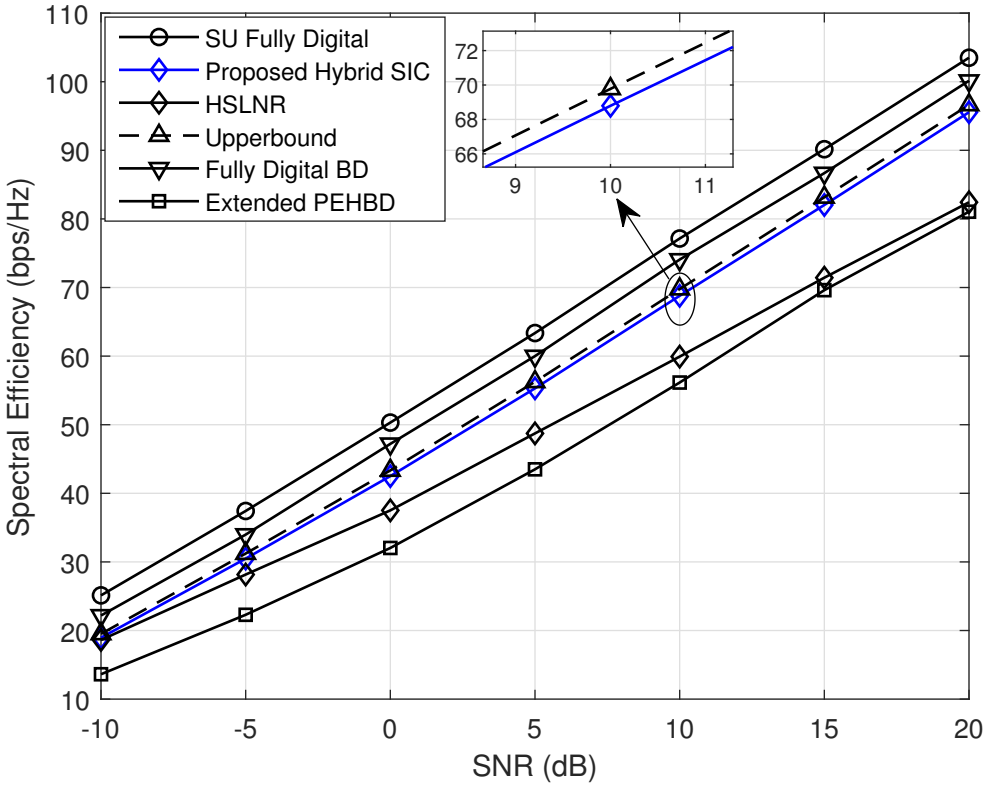

Figure 5.5: Sum SE performance comparison of different algorithms $N_u = 2$, $N_s = 4$ and $N_{RF}^t = 8$, $N_{RF}^r = 4$.

data streams. In Figure 5.9, it is witnessed the sum spectral efficiency improvement achieved by the proposed hybrid OSIC algorithm is about 2%, 4% and 4% compared to the proposed Hybrid SIC algorithm with arbitrary data stream order in the case of $N_u = 8$, $N_u = 4$ and $N_u = 2$, respectively.

5.6.2 Robustness Evaluation

Finally, we illustrate the robustness of the proposed hybrid SIC algorithm in Figure 5.10 and 5.11, adopting the same setup of Figure 5.3. In Figure 5.10, we demonstrate the sum spectral efficiency achieved by the proposed hybrid SIC algorithm with the finite resolution of phase shifters. As observed in Figure 5.10, the performance of the proposed hybrid SIC algorithm for 3-bit, 2-bit and 1-bit resolution phase shifters can achieve over 94%, 90% and 74% performance of infinite resolution phase shifters (referred to as B = ∞), respectively.

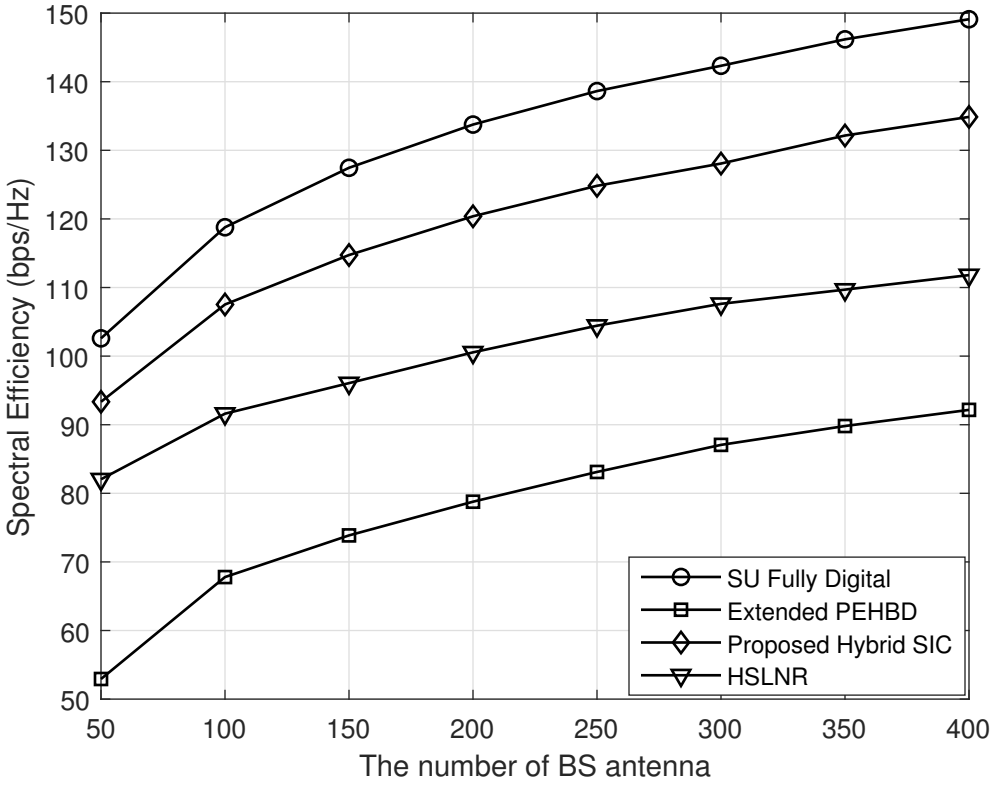

Figure 5.6: Sum SE performance with different number BS antenna, $N_r = 16$, $N_u = 4$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

In Figure 5.11, we evaluate the impact of imperfect CSI on the proposed hybrid SIC beamforming algorithm. From Figure 5.11, we can observe that the performance gap between the proposed hybrid SIC algorithm with imperfect CSI and perfect CSI gets smaller with increasing the reliability of the channel estimate. It is witnessed that the gap is neglectable when $\xi = 0.9$, which confirms our analysis in Section IV.

5.7 Summary

In this chapter, a hybrid beamforming design for wideband mmWave MU-MIMO OFDM systems has been proposed. In our design, we adopt a relaxation of the original mutual information and spectral efficiency maximization problems at the transceiver. For the relaxed problems, we design RF precoder and combiner which can achieve almost the same performance as the original problems. Then a SIC-aided baseband beamforming algorithm on a per-subcarrier basis has been proposed to

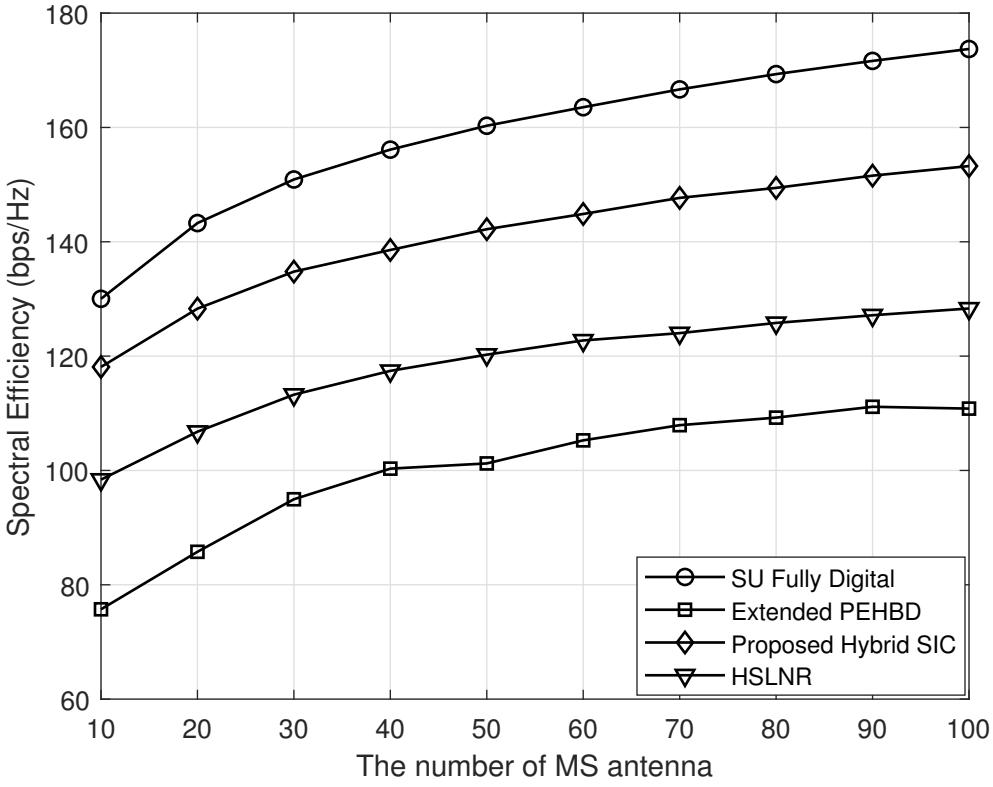

Figure 5.7: Sum SE performance with different number MS antenna, $N_t = 256$, $N_u = 4$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

cancel both inter-user and intra-user interference. The performance of the proposed algorithm in the presence of channel estimation error has been studied. The numerical results demonstrate that the proposed hybrid beamforming algorithm outperforms the existing linear hybrid beamforming at the cost of computational complexity for the SIC procedure. Furthermore, the proposed hybrid beamforming algorithm can achieve good SE performance with finite resolution phase shifters and channel estimation error.

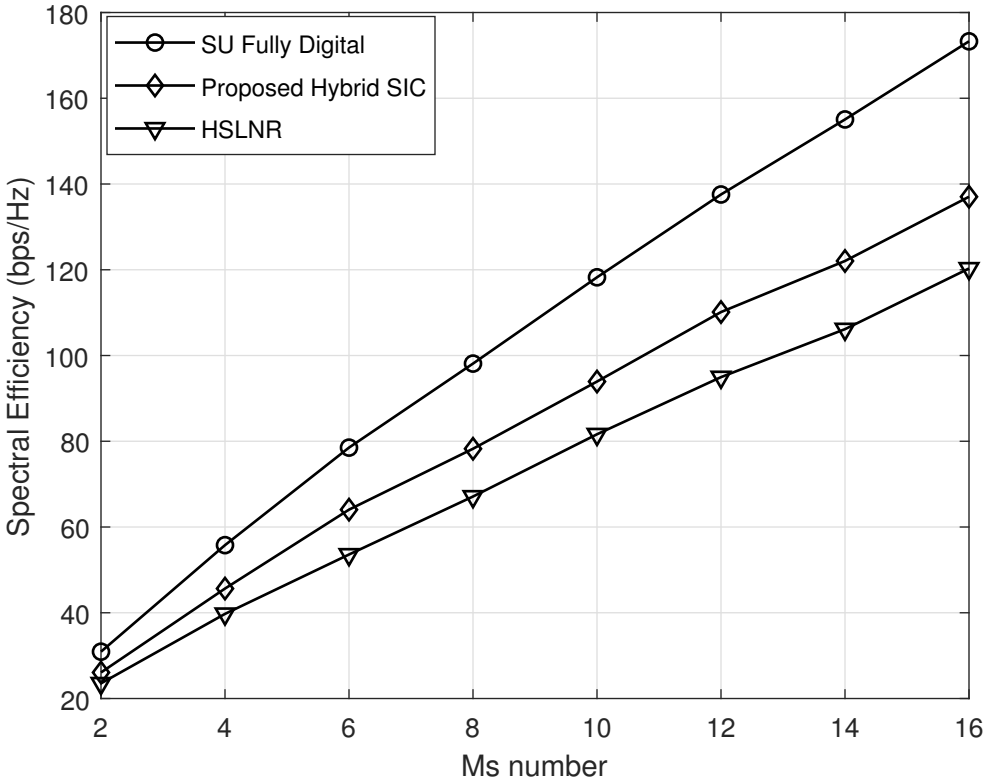

Figure 5.8: Sum SE performance with different MS number, $N_{\mathrm{s}} = 4$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = N_u N_s$.

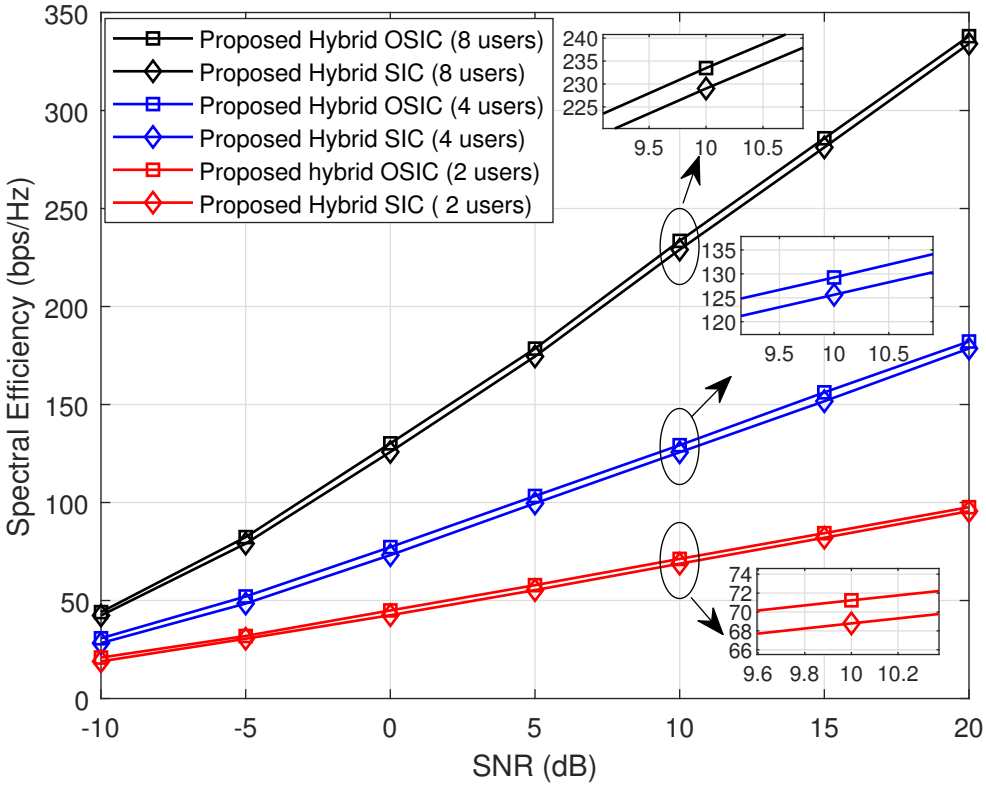

Figure 5.9: Sum SE performance comparison of different algorithms with ordered SIC, $N_u = 8, 4, 2$, $N_{\rm s} = 4$ and $N_{\rm RF}^{\rm t} = N_u N_s$, $N_{\rm RF}^{\rm r} = 4$.

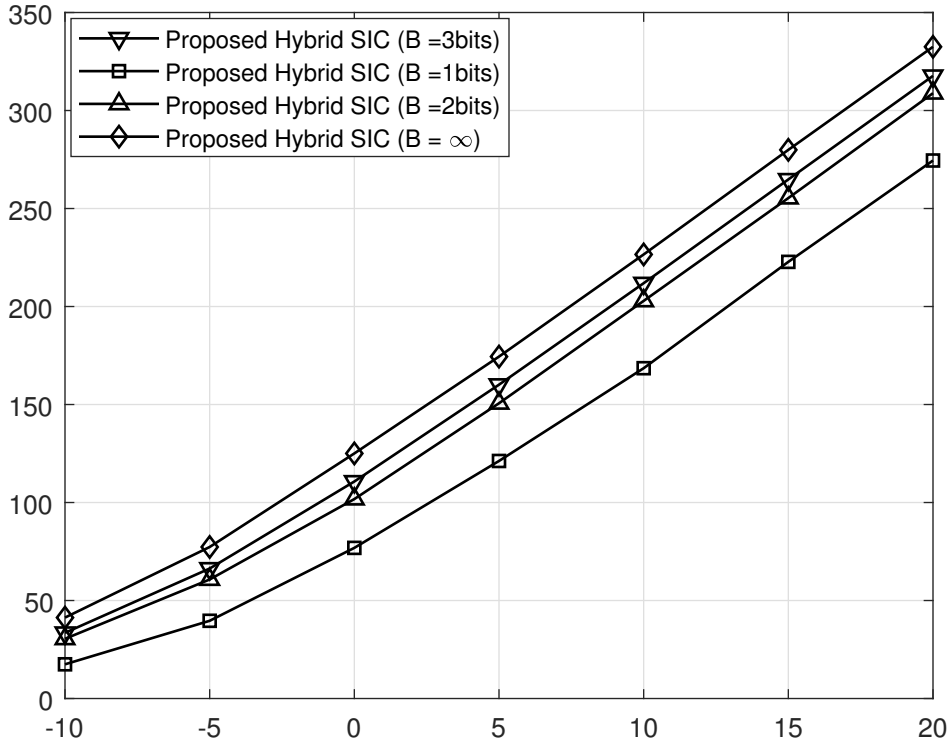

Figure 5.10: Sum SE performance with finite resolution phase shifters, $N_u = 8$, $N_{\mathrm{s}} = 4$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 4$.

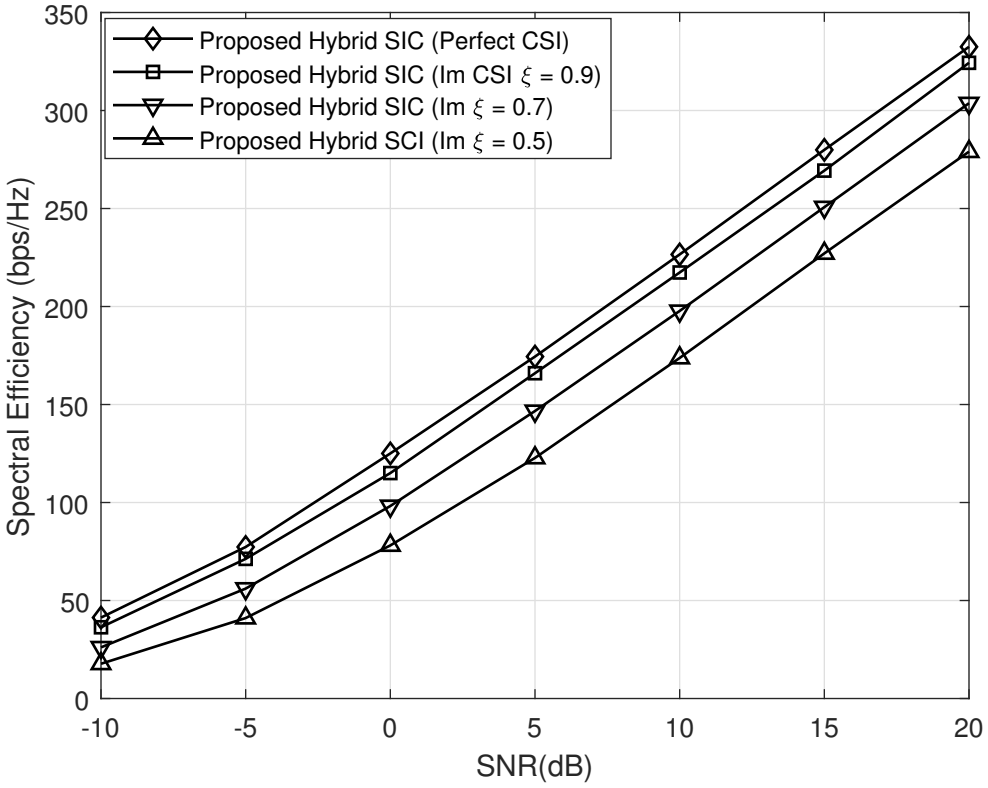

Figure 5.11: Sum SE performance with imperfect CSI, $N_u = 8$, $N_{\mathrm{s}} = 2$ and $N_{\mathrm{RF}}^{\mathrm{t}} = 16$, $N_{\mathrm{RF}}^{\mathrm{r}} = 2$.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we have addressed the hybrid beamforming design for mmWave cellular communications. We have considered both narrowband and wideband mmWave channels. First of all, for narrowband SU-MIMO transmission, we designed a set of orthogonal candidate analog precoder, which can be achieved by performing GSO to the columns of the array response matrix. Array configuration and connection architecture are not required to be considered any more. Furthermore, it can be applied to both fully connected and partially connected architecture. According to the array response vector along which the full digital precoder has the maximum projection, we have proposed a heuristic solution to the order of columns selection to perform GSO.

We then investigate the SIC-aided hybrid precoding and combining design for downlink mmWave multi-user massive MIMO systems, aiming to maximize the overall spectral efficiency. We have proposed an IC-aided hybrid beamforming design framework, in which non-linear algorithms can be used to suppress inter-user and intra-user interference. Under the proposed framework, three SIC-aided hybrid beamforming algorithms have been presented according to the dominance of inter-user interference, intra-user interference, or both. Furthermore, we have proposed the optimal detection order of data streams according to the post-detection SINR, which can further improve the spectral efficiency performance of the proposed three algorithms.

Finally, we have developed a near-optimal closed-form RF beamforming solution for wideband mmWave MU-MIMO OFDM systems. We have reformulated the RF precoder and combiner design problem by a relaxation of the original mutual infor-

mation maximization problem at the transmitter and a relaxation of the spectral efficiency maximization problem at the receiver. For the reformulated problems, we have designed RF precoder and combiner which have almost the same spectral efficiency as the original problems. We have proposed an SIC-aided baseband beamforming algorithm on a per-subcarrier basis to cancel both inter-user and intra-user interference. We have analyzed the performance of the proposed algorithm in the presence of channel estimation error. By modeling the estimation error as independent complex Gaussian random variables and uncorrelated with the ideal channel, tight approximations for both the original mutual information and spectral efficiency have been derived.

6.2 Future Work

It should be noted the hybrid beamforming design proposed in this thesis is based on the assumption that BS has full CSI of all MSs. Future work should be done to investigate efficient channel estimation algorithms, especially in fast fading channels.

As mmWave cellular systems are expected to be densely deployed to guarantee acceptable coverage, spectral efficiency, as well as energy efficiency, inter-cell interference is a major challenge that needs to be solved. It will be interesting to investigate the hybrid beamforming design for multi-cell mmWave cellular systems.

The user fairness is not considered in Chapters 4 and 5. It will also be interesting to investigate how to maximizing the overall system throughput with user fairness.

In addition, the RF beamforming and baseband beamforming proposed in this thesis is designed alternatively. Future work can be focused on how to joint optimization of RF beamformer and baseband beamformer with minimal increase in the computational complexity.

List of Publications

Published papers:

1. Jinlong Zhan, Xiaodai Dong, Yiming Huo, and Yu Zhang. "Gram–Schmidt orthogonalisation aided hybrid precoding in millimetre-wave massive MIMO systems." IET Communications, vol. 14, no. 3 pp. 387-396, October 2019.

2. Jinlong Zhan and Xiaodai Dong, "Interference Cancellation Aided Hybrid Beamforming for mmWave Multi-User Massive MIMO Systems," in IEEE Transactions on Vehicular Technology, vol. 70, no. 3, pp. 2322-2336, March 2021.

3. 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," in IEEE Access, vol. 9, pp. 23398-23408, 2021, doi: 10.1109/ACCESS.2021.3056297.

4. Y. Zhang, Y. Huo, J. Zhan, D. Wang, X. Dong and X. You, "ADMM Enabled Hybrid Precoding in Wideband Distributed Phased Arrays Based MIMO Systems," 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), 2019, pp. 1-5, doi: 10.1109/VTCFall.2019.8891343.

Submitted papers:

Jinlong Zhan and Xiaodai Dong, "Hybrid Beamforming for Wideband Millimeter Wave Massive MIMO Systems," submitted to IEEE Transactions on Wireless Communications.

Appendix A

Proof of Equation (4.28)

From Equation (4.32), after RF combining, the received signal of the k-th MS is given by

$$\begin{aligned}\mathbf{y}_{RF,k} &= \mathbf{W}_{RF,k}^{H}\mathbf{H}_k\mathbf{F}_{RF}\mathbf{F}_{BB,k}^{SO}\mathbf{s}_k \\ &\quad + \mathbf{W}_{RF,k}^{H}\mathbf{H}_k\sum_{l=k+1}^{K}\mathbf{F}_{RF}\mathbf{F}_{BB,l}^{SO}\mathbf{s}_l + \mathbf{W}_{RF,k}^{H}\mathbf{n}_k \\ &= \tilde{\mathbf{H}}_k\mathbf{F}_{BB,k}^{SO}\mathbf{s}_k + \tilde{\mathbf{H}}_k\sum_{l=k+1}^{K}\mathbf{F}_{BB,l}^{SO}\mathbf{s}_l + \mathbf{W}_{RF,k}^{H}\mathbf{n}_k \\ &= \tilde{\mathbf{H}}_k\hat{\mathbf{V}}_k^{(0)}\mathbf{G}_k\mathbf{s}_k + \tilde{\mathbf{H}}_k\sum_{l=k+1}^{K}\mathbf{F}_{BB,l}^{SO}\mathbf{s}_l + \mathbf{W}_{RF,k}^{H}\mathbf{n}_k.\end{aligned} \tag{A.1}$$

The achieved spectral efficiency of the k-th Ms is given by

$$\begin{aligned}R_k &= \log_2\left(\left|\mathbf{I}_{N_{RF}^r} + \mathbf{R}_{I+N}^{-1}\tilde{\mathbf{H}}_k\hat{\mathbf{V}}_k^{(0)}\mathbf{G}_k\mathbf{G}_k^{H}\left(\hat{\mathbf{V}}_k^{(0)}\right)^{H}\tilde{\mathbf{H}}_k^{H}\right|\right) \\ &\overset{(c)}{=} \log_2\left(\left|\mathbf{I}_{N_s} + \mathbf{G}_k^{H}\left(\hat{\mathbf{V}}_k^{(0)}\right)^{H}\tilde{\mathbf{H}}_k^{H}\mathbf{R}_{I+N}^{-1}\tilde{\mathbf{H}}_k\hat{\mathbf{V}}_k^{(0)}\mathbf{G}_k\right|\right),\end{aligned} \tag{A.2}$$

where step (c) of Equation (A.2) is the result of using $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$, $\mathbf{R}_{I+N}$ is the noise and interference covariance matrix, which is given by

$$\mathbf{R}_{I+N} = \sigma^2\mathbf{W}_{RF,k}^{H}\mathbf{W}_{RF,k} + \sum_{l=k+1}^{K}\tilde{\mathbf{H}}_k\mathbf{F}_{BB,l}^{SO}\left(\mathbf{F}_{BB,l}^{SO}\right)^{H}\mathbf{H}_k^{H}. \tag{A.3}$$

The optimal $\mathbf{G}_k$ is the solution to the following optimization problem

$$((\mathbf{G}_k^*)_{k=1:K}) = \arg\max R_k \tag{A.4a}$$

$$\text{s.t.} \quad \text{Tr}\left(\mathbf{G}_k\mathbf{G}_k^H\right) \leq P_k, \tag{A.4b}$$

where P_k is the transmit power of the k-th MS. Finding $\mathbf{G}_k$ to maximize the achieved spectral efficiency leads to a water-filling solution using the following SVD

$$\begin{aligned}\left(\hat{\mathbf{V}}_k^{(0)}\right)^H \tilde{\mathbf{H}}_k^H \mathbf{R}_{I+N}^{-1} \tilde{\mathbf{H}}_k \hat{\mathbf{V}}_k^{(0)} &= \mathbf{M}_k \mathbf{\Lambda}_k \mathbf{M}_k^H \\ &= \mathbf{M}_k \mathbf{\Lambda}_k \left[\mathbf{M}_k^{(N_s)}, \mathbf{M}_k^{\left(N_{RF}^r - N_s\right)}\right]^H,\end{aligned} \tag{A.5}$$

where $\mathbf{M}_k^{(N_s)}$ consists of the first N_s columns of $\mathbf{M}_k$, and $\mathbf{M}_k^{\left(N_{RF}^r - N_s\right)}$holds the rest $N_{RF}^r - N_s$ columns of $\mathbf{M}_k$. The optimal solution of $\mathbf{G}_k$, which is given by

$$\mathbf{G}_k = \mathbf{M}_k^{(N_s)} \mathbf{\Gamma}_k^{1/2}, \tag{A.6}$$

where the water-filling coefficients in $\mathbf{\Gamma}_k$ are chosen such that maximum spectral efficiency is achieved. The total transmit power for all users is then the sum of the elements of all $\mathbf{\Gamma}_k$. Therefore, the baseband precoder of the k-th MS is obtained

$$\mathbf{F}_{BB,k}^{SO} = \hat{\mathbf{V}}_k^{(0)} \mathbf{G}_k = \hat{\mathbf{V}}_k^{(0)} \mathbf{M}_k^{(N_s)} \mathbf{\Gamma}_k^{1/2}. \tag{A.7}$$

Appendix B

Solution to Problem (5.16)

From Equation (5.16), the Frobenius norm between the optimal unconstrained RF precoder and constrained precoder can be rewritten as

$$\begin{aligned}
&\left\|\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right\|_F^2 \\
&=\mathrm{Tr}\left\{\left(\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right)\left(\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right)^H\right\} \\
&=\mathrm{Tr}\left\{\left(\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right)^H\left(\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right)\right\} \\
&\overset{(e)}{=}2N_{RF}^t-\mathrm{Tr}\left\{2\mathrm{Re}\left\{\mathbf{F}_{RF}^H\mathbf{F}_{RFun}\right\}\right\}.
\end{aligned} \tag{B.1}$$

The simplification step (e) of Equation (B.1) holds due to $\mathbf{F}_{RFun}^H\mathbf{F}_{RFun}=\mathbf{I}_{N_{RF}^t}$ and $\mathbf{F}_{RF}^H\mathbf{F}_{RF}=\mathbf{I}_{N_{RF}^t}$ [96]. $\mathbf{F}_{RF}^H\mathbf{F}_{RFun}$ can be further expressed as

$$\begin{aligned}
&\mathbf{F}_{RF}^H\mathbf{F}_{RFun} \\
&=\sum_{i=1}^{N_{RF}^t}\sum_{j=1}^{N_t}\left|\mathbf{F}_{RF}^*\left(i,j\right)\right|\left|\mathbf{F}_{RFun}\left(i,j\right)\right|e^{j\phi(i,j)},
\end{aligned} \tag{B.2}$$

where $\phi\left(i,j\right)=\angle\mathbf{F}_{RF}^*\left(i,j\right)-\angle\mathbf{F}_{RFun}\left(i,j\right)$. It can be seen from Equation (B.2) that when $\phi\left(i,j\right)=0$, $\left\|\mathbf{F}_{RFun}-\mathbf{F}_{RF}\right\|_F^2$ will be minimized. Therefore, $\mathbf{F}_{RF}$ can be expressed as

$$\mathbf{F}_{RF}=\frac{1}{\sqrt{N_t}}e^{j\angle\mathbf{F}_{RFun}}=\frac{1}{\sqrt{N_t}}e^{j\angle\mathbf{V}_{ave}\left(:,1:N_{RF}^t\right)}, \tag{B.3}$$

which completes the proof.

Appendix C

Proof of Equation (5.28)

From Equation (5.18), after RF combining, the received signal of the i-th MS at subcarrier k is given by

$$\begin{aligned}
\mathbf{y}_{RF,i}[k] &= \mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\mathbf{F}_{RF}\mathbf{F}_{BB,i}^{SIC}[k]\mathbf{s}_i[k] \\
&\quad + \mathbf{W}_{RF,i}^{H}\mathbf{H}_i[k]\sum_{j=i+1}^{N_u}\mathbf{F}_{RF}\mathbf{F}_{BB,j}^{SIC}[k]\mathbf{s}_j[k] \\
&\quad + \mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k] \\
&= \tilde{\mathbf{H}}_i[k]\mathbf{F}_{BB,i}^{SIC}[k]\mathbf{s}_i[k] + \tilde{\mathbf{H}}_i[k]\sum_{j=i+1}^{N_u}\mathbf{F}_{BB,j}^{SIC}[k]\mathbf{s}_j[k] \\
&\quad + \mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k] \\
&= \tilde{\mathbf{H}}_i[k]\hat{\mathbf{V}}_i^{(0)[k]}\mathbf{G}_i[k]\mathbf{s}_i[k] + \tilde{\mathbf{H}}_i[k]\sum_{j=i+1}^{N_u}\mathbf{F}_{BB,j}^{SIC}[k]\mathbf{s}_j[k] \\
&\quad + \mathbf{W}_{RF,i}^{H}\mathbf{n}_i[k].
\end{aligned} \tag{C.1}$$

The achieved spectral efficiency of the i-th Ms at subcarrier k is given by

$$\begin{aligned}
R_i[k] &= \log_2\left(\left|\mathbf{I}_{N_{RF}^r} + \mathbf{R}_{I+N}^{-1}[k]\tilde{\mathbf{H}}_i[k]\hat{\mathbf{V}}_i^{(0)}[k]\mathbf{G}_i[k]\mathbf{G}_i^{H}[k]\left(\hat{\mathbf{V}}_i^{(0)}[k]\right)^{H}\tilde{\mathbf{H}}_i^{H}[k]\right|\right) \\
&\overset{(g)}{=} \log_2\left(\left|\mathbf{I}_{N_s} + \mathbf{G}_i^{H}[k]\left(\hat{\mathbf{V}}_i^{(0)}[k]\right)^{H}\tilde{\mathbf{H}}_i^{H}[k]\mathbf{R}_{I+N}^{-1}[k]\tilde{\mathbf{H}}_i[k]\hat{\mathbf{V}}_i^{(0)}[k]\mathbf{G}_i[k]\right|\right)
\end{aligned} \tag{C.2}$$

where step (g) of Equation (C.2) is the result of using $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$, $\mathbf{R}_{I+N}$ is the noise and interference covariance matrix, which is given by

$$\begin{aligned}
\mathbf{R}_{I+N}[k] &= \sigma^2\mathbf{W}_{RF,i}^{H}\mathbf{W}_{RF,i} \\
&\quad + \sum_{j=i+1}^{N_u}\tilde{\mathbf{H}}_i[k]\mathbf{F}_{BB,j}^{SIC}[k]\left(\mathbf{F}_{BB,j}^{SIC}[k]\right)^{H}\mathbf{H}_i^{H}[k].
\end{aligned} \tag{C.3}$$

The optimal $\mathbf{G}_i[k]$ is the solution to the following optimization problem

$$\left((\mathbf{G}_i^*[k])_{i=1:N_u}\right) = \arg\max R_i[k] \tag{C.4a}$$

$$\text{s.t.} \quad Tr\left(\mathbf{G}_i[k]\mathbf{G}_i^H[k]\right) \leq P_i, \tag{C.4b}$$

where P_i is the transmit power of the i-th MS. Finding $\mathbf{G}_i[k]$ to maximize the achieved spectral efficiency leads to a water-filling solution using the following SVD

$$\begin{aligned}\left(\hat{\mathbf{V}}_i^{(0)}[k]\right)^H \tilde{\mathbf{H}}_i^H[k]\mathbf{R}_{I+N}^{-1}[k]\tilde{\mathbf{H}}_i[k]\hat{\mathbf{V}}_i^{(0)[k]} \\ = \mathbf{M}_i[k]\mathbf{\Lambda}_i[k]\mathbf{M}_i^H[k] \\ = \mathbf{M}_i[k]\mathbf{\Lambda}_i[k]\left[\mathbf{M}_i^{(N_s)}[k], \mathbf{M}_i^{\left(N_{RF}^r - N_s\right)}[k]\right]^H\end{aligned} \tag{C.5}$$

where $\mathbf{M}_i^{(N_s)}[k]$ consists of the first N_s columns of $\mathbf{M}_i[k]$, and $\mathbf{M}_i^{\left(N_{RF}^r - N_s\right)[k]}$holds the rest $N_{RF}^r - N_s$ columns of $\mathbf{M}_i[k]$. The optimal solution of $\mathbf{G}_i[k]$, which is given by

$$\mathbf{G}_i[k] = \mathbf{M}_i^{(N_s[k])}\mathbf{\Gamma}_i^{1/2}[k], \tag{C.6}$$

where the water-filling coefficients in $\mathbf{\Gamma}_k$ are chosen such that maximum spectral efficiency is achieved. The total transmit power for all users is then the sum of the elements of all $\mathbf{\Gamma}_i[k]$. Therefore, the baseband precoder of the i-th MS at subcarrier k is obtained

$$\mathbf{F}_{BB,i}^{SIC}[k] = \hat{\mathbf{V}}_i^{(0)}[k]\mathbf{G}_i[k] = \hat{\mathbf{V}}_i^{(0)}[k]\mathbf{M}_i^{(N_s[k])}\mathbf{\Gamma}_i^{1/2}[k]. \tag{C.7}$$

Bibliography

[1] Zahid Ghadialy. An introduction to 5g and 'real' 5g. https://www.slideshare.net/3G4GLtd/an-introduction-to-5g-and-real-5g, 2020.

[2] NTIA. United states frequency allocation chart. https://www.ntia.doc.gov/files/ntia/publications/january_2016_spectrum_wall_chart.pdf, 2016.

[3] Gabriele Falciasecca. Marconi's early experiments in wireless telegraphy, 1895. *IEEE Antennas and Propagation Magazine*, 52(6):220–221, 2010.

[4] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[5] DataReportal. Digital around the world. https://datareportal.com/global-digital-overview, 2021.

[6] Theodore S Rappaport et al. *Wireless communications: principles and practice*, volume 2. prentice hall PTR New Jersey, 1996.

[7] Andreas F Molisch. *Wireless communications*, volume 34. John Wiley & Sons, 2012.

[8] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[9] Richard Galazzo. From 1g to 5g: The history of cell phones and their cellular generations. https://www.cengn.ca/information-centre/innovation/timeline-from-1g-to-5g-a-brief-history-on-cell-phones/, 2021.

[10] Reinhardt Haverans. From 1g to 5g: A brief history of the evolution of mobile standards. https://www.brainbridge.be/en/blog/1g-5g-brief-history-evolution-mobile-standards, 2021.

[11] Wei Hong, Zhi Hao Jiang, Chao Yu, Debin Hou, Haiming Wang, Chong Guo, Yun Hu, Le Kuai, Yingrui Yu, Zhengbo Jiang, et al. The role of millimeter-wave technologies in 5g/6g wireless communications. *IEEE Journal of Microwaves*, 1(1):101–122, 2021.

[12] Walid Saad, Mehdi Bennis, and Mingzhe Chen. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems. *IEEE network*, 34(3):134–142, 2020.

[13] Shilpa Talwar, Nageen Himayat, Hosein Nikopour, Feng Xue, Geng Wu, and Vida Ilderem. 6g: Connectivity in the era of distributed intelligence. *IEEE Communications Magazine*, 59(11):45–50, 2021.

[14] Zhi Chen, Chong Han, Yongzhi Wu, Lingxiang Li, Chongwen Huang, Zhaoyang Zhang, Guangjian Wang, and Wen Tong. Terahertz wireless communications for 2030 and beyond: A cutting-edge frontier. *IEEE Communications Magazine*, 59(11):66–72, 2021.

[15] 3GPP. NR; Base Station (BS) radio transmission and reception. Technical Specification (TS) 36.104, 3rd Generation Partnership Project (3GPP), 03 2017. Version 14.2.2.

[16] Thomas L Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wireless Commun.*, 9(11):3590–3600, Nov. 2010.

[17] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson. Scaling up mimo: opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, 30(1):40–60, 2013.

[18] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave mimo systems. *IEEE J. Special Topics Signal Process.*, 10(3):436–453, Apr. 2016.

[19] A Lee Swindlehurst, Ender Ayanoglu, Payam Heydari, and Filippo Capolino. Millimeter-wave massive mimo: The next wireless revolution? *IEEE Commun. Mag.*, 52(9):56–62, Sep. 2014.

[20] Shuangfeng Han, I Chih-Lin, Zhikun Xu, and Corbett Rowell. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun. Mag.*, 53(1):186–194, Jan. 2015.

[21] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results. *IEEE Commun. Mag.*, 52(2):106–113, Feb. 2014.

[22] Theodore S Rappaport, Robert W Heath Jr, Robert C Daniels, and James N Murdock. *Millimeter wave wireless communications*. Pearson Education, 2015.

[23] Cheng-Xiang Wang, Fourat Haider, Xiqi Gao, Xiao-Hu You, Yang Yang, Dongfeng Yuan, Hadi M Aggoune, Harald Haas, Simon Fletcher, and Erol Hepsaydir. Cellular architecture and key technologies for 5g wireless communication networks. *IEEE communications magazine*, 52(2):122–130, 2014.

[24] Kiran Venugopal, Matthew C Valenti, and Robert W Heath. Device-to-device millimeter wave communications: Interference, coverage, rate, and finite topologies. *IEEE Transactions on Wireless Communications*, 15(9):6175–6188, 2016.

[25] Leila Tlebaldiyeva, Behrouz Maham, and Theodoros A Tsiftsis. Capacity analysis of device-to-device mmwave networks under transceiver distortion noise and imperfect csi. *IEEE Transactions on Vehicular Technology*, 69(5):5707–5712, 2020.

[26] Junil Choi, Vutha Va, Nuria Gonzalez-Prelcic, Robert Daniels, Chandra R Bhat, and Robert W Heath. Millimeter-wave vehicular communication to support massive automotive sensing. *IEEE Communications Magazine*, 54(12):160–167, 2016.

[27] Linghe Kong, Muhammad Khurram Khan, Fan Wu, Guihai Chen, and Peng Zeng. Millimeter-wave wireless communications for iot-cloud supported autonomous vehicles: Overview, design, and challenges. *IEEE Communications Magazine*, 55(1):62–68, 2017.

[28] Francisco J Martin-Vega, Mari Carmen Aguayo-Torres, Gerardo Gomez, Jose Tomas Entrambasaguas, and Trung Q Duong. Key technologies, model-

ing approaches, and challenges for millimeter-wave vehicular communications. *IEEE Communications Magazine*, 56(10):28–35, 2018.

[29] Junyi Wang, Zhou Lan, Chang-woo Pyo, Tuncer Baykas, Chin-sean Sum, Mohammad Azizur Rahman, Jing Gao, Ryuhei Funada, Fumihide Kojima, Hiroshi Harada, et al. Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems. *IEEE Journal on Selected Areas in Communications*, 27(8):1390–1399, 2009.

[30] Sooyoung Hur, Taejoon Kim, David J Love, James V Krogmeier, Timothy A Thomas, and Amitava Ghosh. Millimeter wave beamforming for wireless backhaul and access in small cell networks. *IEEE transactions on communications*, 61(10):4391–4403, 2013.

[31] Roi Méndez-Rial, Cristian Rusu, Nuria González-Prelcic, Ahmed Alkhateeb, and Robert W Heath. Hybrid mimo architectures for millimeter wave communications: Phase shifters or switches? *Ieee Access*, 4:247–267, 2016.

[32] Qinghua Li, Guangjie Li, Wookbong Lee, Moon-il Lee, David Mazzarese, Bruno Clerckx, and Zexian Li. Mimo techniques in wimax and lte: a feature overview. *IEEE Communications magazine*, 48(5):86–92, 2010.

[33] Ahmed Alkhateeb, Jianhua Mo, Nuria Gonzalez-Prelcic, and Robert W Heath. Mimo precoding and combining solutions for millimeter-wave systems. *IEEE Communications Magazine*, 52(12):122–131, 2014.

[34] Le Kuai, Jixin Chen, Zhi Hao Jiang, Chao Yu, Chong Guo, Yingrui Yu, Hou-Xing Zhou, and Wei Hong. A n260 band 64 channel millimeter wave full-digital multi-beam array for 5g massive mimo applications. *IEEE Access*, 8:47640–47653, 2020.

[35] Yiming Huo, Xiaodai Dong, and Wei Xu. 5G cellular user equipment: From theory to practical hardware design. *IEEE Access*, 5:13992–14010, Aug. 2017.

[36] Yiming Huo, Xiaodai Dong, Wei Xu, and Marvin Yuen. Enabling multi-functional 5g and beyond user equipment: A survey and tutorial. *IEEE Access*, 7:116975–117008, 2019.

[37] Chinh H Doan, Sohrab Emami, David A Sobel, Ali M Niknejad, and Robert W Brodersen. Design considerations for 60 ghz cmos radios. *IEEE Communications Magazine*, 42(12):132–140, 2004.

[38] Jian A Zhang, Xiaojing Huang, Val Dyadyuk, and Y Jay Guo. Massive hybrid antenna array for millimeter-wave cellular communications. *IEEE Wireless Communications*, 22(1):79–87, 2015.

[39] Theodore S Rappaport, James N Murdock, and Felix Gutierrez. State of the art in 60-ghz integrated circuits and systems for wireless communications. *Proceedings of the IEEE*, 99(8):1390–1436, 2011.

[40] Bin Le, Thomas W Rondeau, Jeffrey H Reed, and Charles W Bostian. Analog-to-digital converters. *IEEE Signal Processing Magazine*, 22(6):69–77, 2005.

[41] Brian A Floyd, Scott K Reynolds, Ullrich R Pfeiffer, Thomas Zwick, Troy Beukema, and Brian Gaucher. Sige bipolar transceiver circuits operating at 60 ghz. *IEEE journal of solid-state circuits*, 40(1):156–167, 2005.

[42] Tim LaRocca, Jenny Yi-Chun Liu, and Mau-Chung Frank Chang. 60 ghz cmos amplifiers using transformer-coupling and artificial dielectric differential transmission lines for compact design. *IEEE journal of solid-state circuits*, 44(5):1425–1435, 2009.

[43] Terry Yao, Michael Q Gordon, Keith KW Tang, Kenneth HK Yau, Ming-Ta Yang, Peter Schvan, and Sorin P Voinigescu. Algorithmic design of cmos lnas and pas for 60-ghz radio. *IEEE Journal of Solid-State Circuits*, 42(5):1044–1057, 2007.

[44] Yikun Yu, Peter GM Baltus, Anton de Graauw, Edwin van der Heijden, Cicero S Vaucher, and Arthur HM van Roermund. A 60 ghz phase shifter integrated with lna and pa in 65 nm cmos for phased array systems. *IEEE Journal of Solid-State Circuits*, 45(9):1697–1709, 2010.

[45] Arun Natarajan, Scott K Reynolds, Ming-Da Tsai, Sean T Nicolson, Jing-Hong Conan Zhan, Dong Gun Kam, Duixian Liu, Yen-Lin Oscar Huang, Alberto Valdes-Garcia, and Brian A Floyd. A fully-integrated 16-element phased-array receiver in sige bicmos for 60-ghz communications. *IEEE journal of solid-state circuits*, 46(5):1059–1075, 2011.

[46] Sang Young Kim, Dong-Woo Kang, Kwang-Jin Koh, and Gabriel M Rebeiz. An improved wideband all-pass i/q network for millimeter-wave phase shifters. *IEEE transactions on microwave theory and techniques*, 60(11):3431–3439, 2012.

[47] Jing-Lin Kuo, Yi-Fong Lu, Ting-Yi Huang, Yi-Long Chang, Yi-Keng Hsieh, Pen-Jui Peng, I-Chih Chang, Tzung-Chuen Tsai, Kun-Yao Kao, Wei-Yuan Hsiung, et al. 60-ghz four-element phased-array transmit/receive system-in-package using phase compensation techniques in 65-nm flip-chip cmos process. *IEEE Transactions on Microwave Theory and Techniques*, 60(3):743–756, 2012.

[48] Wei-Tsung Li, Yun-Chieh Chiang, Jeng-Han Tsai, Hong-Yuan Yang, Jen-Hao Cheng, and Tian-Wei Huang. 60-ghz 5-bit phase shifter with integrated vga phase-error compensation. *IEEE Transactions on Microwave Theory and Techniques*, 61(3):1224–1235, 2013.

[49] Pradeep Shettigar and Shanthi Pavan. A 15mw 3.6 gs/s ct-$\delta\sigma$ adc with 36mhz bandwidth and 83db dr in 90nm cmos. In *2012 IEEE International Solid-State Circuits Conference*, pages 156–158. IEEE, 2012.

[50] Dong-Shin Jo, Il-Hoon Jang, Dong-Suk Lee, Yong-Sang You, Yong-Hee Lee, Ho-Jin Park, Seung-Tak Ryu, et al. 26.4 a 21fj/conv-step 9 enob 1.6 gs/s 2× time-interleaved fati sar adc with background offset and timing-skew calibration in 45nm cmos. In *2015 IEEE International Solid-State Circuits Conference-(ISSCC) Digest of Technical Papers*, pages 1–3. IEEE, 2015.

[51] Nicolas Le Dortz, Jean-Pierre Blanc, Thierry Simon, Sarah Verhaeren, Emmanuel Rouat, Pascal Urard, Stéphane Le Tual, Dimitri Goguet, Caroline Lelandais-Perrault, and Philippe Benabes. 22.5 a 1.62 gs/s time-interleaved sar adc with digital background mismatch calibration achieving interleaving spurs below 70dbfs. In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pages 386–388. IEEE, 2014.

[52] Erwin Janssen, Kostas Doris, Athon Zanikopoulos, Alessandro Murroni, Gerard Van der Weide, Yu Lin, Ludo Alvado, Frederic Darthenay, and Yannick Fregeais. An 11b 3.6 gs/s time-interleaved sar adc in 65nm cmos. In *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pages 464–465. IEEE, 2013.

[53] Xinying Zhang, Andreas F Molisch, and Sun-Yuan Kung. Variable-phase-shift-based RF-baseband codesign for mimo antenna selection. *IEEE Trans. Signal Process.*, 53(11):4091–4103, Nov. 2005.

[54] Vijay Venkateswaran and Alle-Jan van der Veen. Analog beamforming in mimo communications with phase shift networks and online channel estimation. *IEEE Trans. Signal Process.*, 58(8):4131–4143, Aug. 2010.

[55] Pallav Sudarshan, Neelesh B Mehta, Andreas F Molisch, and Jin Zhang. Channel statistics-based rf pre-processing with antenna selection. *IEEE Trans. Wireless Commun.*, 5(12):3501–3511, Dec. 2006.

[56] Foad Sohrabi and Wei Yu. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE J. Special Topics Signal Process.*, 10(3):501–513, Apr. 2016.

[57] Tadilo Endeshaw Bogale, Long Bao Le, Afshin Haghighat, and Luc Vandendorpe. On the number of rf chains and phase shifters, and scheduling design with hybrid analog–digital beamforming. *IEEE Transactions on Wireless Communications*, 15(5):3311–3326, 2016.

[58] Omar El Ayach, Sridhar Rajagopal, Shadi Abu-Surra, Zhouyue Pi, and Robert W Heath. Spatially sparse precoding in millimeter wave mimo systems. *IEEE Trans. Wireless Commun.*, 13(3):1499–1513, Mar. 2014.

[59] Weiheng Ni and Xiaodai Dong. Hybrid block diagonalization for massive multiuser mimo systems. *IEEE Trans. Commun.*, 64(1):201–211, Jan. 2016.

[60] Le Liang, Wei Xu, and Xiaodai Dong. Low-complexity hybrid precoding in massive multiuser mimo systems. *IEEE Wireless Commun. Lett.*, 3(6):653–656, Dec. 2014.

[61] Ahmed Alkhateeb, Geert Leus, and Robert W Heath. Limited feedback hybrid precoding for multi-user millimeter wave systems. *IEEE transactions on wireless communications*, 14(11):6481–6494, 2015.

[62] Xiaoyong Wu, Danpu Liu, and Fangfang Yin. Hybrid beamforming for multi-user massive mimo systems. *IEEE Transactions on Communications*, 66(9):3879–3891, 2018.

[63] Xianghao Yu, Juei-Chin Shen, Jun Zhang, and Khaled B Letaief. Alternating minimization algorithms for hybrid precoding in millimeter wave mimo systems. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):485–500, 2016.

[64] Xinyu Gao, Linglong Dai, Shuangfeng Han, I Chih-Lin, and Robert W Heath. Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays. *IEEE J. Sel. Areas in Commun.*, 34(4):998–1009, Apr. 2016.

[65] Jinlong Zhan and Xiaodai Dong. Interference cancellation aided hybrid beamforming for mmwave multi-user massive mimo systems. *IEEE Transactions on Vehicular Technology*, 70(3):2322–2336, 2021.

[66] Hongji Huang, Yiwei Song, Jie Yang, Guan Gui, and Fumiyuki Adachi. Deep-learning-based millimeter-wave massive mimo for hybrid precoding. *IEEE Transactions on Vehicular Technology*, 68(3):3027–3032, 2019.

[67] Ahmet M Elbir and Anastasios K Papazafeiropoulos. Hybrid precoding for multiuser millimeter wave massive mimo systems: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 69(1):552–563, 2019.

[68] Wenyan Ma, Chenhao Qi, Zaichen Zhang, and Julian Cheng. Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive mimo. *IEEE Transactions on Communications*, 68(5):2838–2849, 2020.

[69] Mohammed A Almagboul, Feng Shu, and Abdeldime MS Abdelgader. Deep-learning based phase-only robust massive mu-mimo hybrid beamforming. *IEEE Communications Letters*, 25(7):2280–2284, 2021.

[70] Hamed Hojatian, Jérémy Nadal, Jean-François Frigon, and François Leduc-Primeau. Unsupervised deep learning for massive mimo hybrid beamforming. *IEEE Transactions on Wireless Communications*, pages 1–1, 2021.

[71] Shu Sun, Theodore S Rappaport, Mansoor Shafi, Pan Tang, Jianhua Zhang, and Peter J Smith. Propagation models and performance evaluation for 5g millimeter-wave bands. *IEEE Transactions on Vehicular Technology*, 67(9):8422–8439, 2018.

[72] Zhouyue Pi and Farooq Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE communications magazine*, 49(6):101–107, 2011.

[73] Hang Zhao, Rimma Mayzus, Shu Sun, Mathew Samimi, Jocelyn K Schulz, Yaniv Azar, Kevin Wang, George N Wong, Felix Gutierrez, and Theodore S Rappaport. 28 ghz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in new york city. In *2013 IEEE international conference on communications (ICC)*, pages 5163–5167. IEEE, 2013.

[74] Jonathan S Lu, Daniel Steinbach, Patrick Cabrol, and Philip Pietraski. Modeling human blockers in millimeter wave radio links. *ZTE communications*, 10(4):23–28, 2012.

[75] Felix K Schwering, Edmond J Violette, and Richard H Espeland. Millimeter-wave propagation in vegetation: Experiments and theory. *IEEE Transactions on Geoscience and Remote Sensing*, 26(3):355–367, 1988.

[76] Gary M Comparetto. Impact of dust and foliage on signal attenuation in the millimeter wave regime. In *Atmospheric Propagation and Remote Sensing II*, volume 1968, pages 81–94. International Society for Optics and Photonics, 1993.

[77] Theodore S Rappaport, Yijun Qiao, Jonathan I Tamir, James N Murdock, and Eshar Ben-Dor. Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems. In *2012 IEEE Radio and Wireless Symposium*, pages 151–154. IEEE, 2012.

[78] Theodore S Rappaport, Felix Gutierrez, Eshar Ben-Dor, James N Murdock, Yijun Qiao, and Jonathan I Tamir. Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. *IEEE transactions on antennas and propagation*, 61(4):1850–1859, 2012.

[79] Mustafa Riza Akdeniz, Yuanpeng Liu, Mathew K Samimi, Shu Sun, Sundeep Rangan, Theodore S Rappaport, and Elza Erkip. Millimeter wave channel modeling and cellular capacity evaluation. *IEEE journal on selected areas in communications*, 32(6):1164–1179, 2014.

[80] Mathew K Samimi and Theodore S Rappaport. 3-d millimeter-wave statistical channel model for 5g wireless system design. *IEEE Transactions on Microwave Theory and Techniques*, 64(7):2207–2225, 2016.

[81] Theodore S Rappaport, George R MacCartney, Shu Sun, Hangsong Yan, and Sijia Deng. Small-scale, local area, and transitional millimeter wave propagation for 5g communications. *IEEE Transactions on Antennas and Propagation*, 65(12):6474–6490, 2017.

[82] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. Millimeter wave mobile communications for 5g cellular: It will work! *IEEE access*, 1:335–349, 2013.

[83] Hao Xu, Vikas Kukshya, and Theodore S Rappaport. Spatial and temporal characteristics of 60-ghz indoor channels. *IEEE Journal on selected areas in communications*, 20(3):620–630, 2002.

[84] A Balanis Constantine. *Antenna Theory*. Wiley, 1997.

[85] Ahmed Alkhateeb and Robert W Heath. Frequency selective hybrid precoding for limited feedback millimeter wave systems. *IEEE Trans. Commun.*, 64(5):1801–1818, May 2016.

[86] Sungwoo Park, Ahmed Alkhateeb, and Robert W Heath. Dynamic subarrays for hybrid precoding in wideband mmwave mimo systems. *IEEE Transactions on Wireless Communications*, 16(5):2907–2920, 2017.

[87] Wei-Lun Hung, Chiang-Hen Chen, Ching-Chun Liao, Cheng-Rung Tsai, and An-Yeu Andy Wu. Low-complexity hybrid precoding algorithm based on orthogonal beamforming codebook. In *Proc. IEEE Workshop on Signal Processing Systems (SiPS)*, pages 1–5, Hangzhou, China, Oct. 2015.

[88] Cristian Rusu, Roi Méndez-Rial, Nuria González-Prelcicy, and Robert W Heath. Low complexity hybrid sparse precoding and combining in millimeter wave mimo systems. In *Proc. IEEE International Conference on Communications (ICC)*, pages 1340–1345, London, UK, Jun. 2015.

[89] Cristian Rusu, Roi Mendez-Rial, Nuria González-Prelcic, and Robert W Heath. Low complexity hybrid precoding strategies for millimeter wave communication systems. *IEEE Trans. Wireless Commun.*, 15(12):8380–8393, Dec. 2016.

[90] Hussein Seleem, Ahmed Iyanda Sulyman, and Abdulhameed Alsanie. Hybrid precoding-beamforming design with hadamard rf codebook for mmwave large-scale mimo systems. *IEEE Access*, 5:6813–6823, Jun. 2017.

[91] Weiheng Ni, Xiaodai Dong, and Wu-Sheng Lu. Near-optimal hybrid processing for massive mimo systems via matrix decomposition. *IEEE transactions on signal processing*, 65(15):3922–3933, 2017.

[92] Juening Jin, Yahong Rosa Zheng, Wen Chen, and Chengshan Xiao. Hybrid precoding for millimeter wave mimo systems: A matrix factorization approach. *IEEE Transactions on Wireless Communications*, 17(5):3327–3339, 2018.

[93] Chengwen Xing, Xin Zhao, Wei Xu, Xiaodai Dong, and Geoffrey Ye Li. A framework on hybrid mimo transceiver design based on matrix-monotonic optimization. *IEEE Transactions on Signal Processing*, 2019.

[94] Tian Lin, Jiaqi Cong, Yu Zhu, Jun Zhang, and Khaled Ben Letaief. Hybrid beamforming for millimeter wave systems using the mmse criterion. *IEEE Transactions on Communications*, 67(5):3693–3708, 2019.

[95] Shahar Stein Ioushua and Yonina C Eldar. A family of hybrid analog–digital beamforming methods for massive mimo systems. *IEEE Transactions on Signal Processing*, 67(12):3243–3257, 2019.

[96] Foad Sohrabi and Wei Yu. Hybrid analog and digital beamforming for mmwave ofdm large-scale antenna arrays. *IEEE Journal on Selected Areas in Communications*, 35(7):1432–1443, 2017.

[97] Steven Roman, S Axler, and FW Gehring. *Advanced Linear Algebra*. Springer, 2005.

[98] Ahmed Alkhateeb and Robert W Heath. Gram schmidt based greedy hybrid precoding for frequency selective millimeter wave mimo systems. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3396–3400, Shanghai, China, Mar. 2016.

[99] Wenjuan Pu, Xiaohui Li, Yingchao Lin, and Ruiyang Yuan. Low complexity hybrid beamforming based on orthogonal constraint and phase extraction. In *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–5, San Francisco, CA, Mar. 2017.

[100] Duy HN Nguyen, Long Bao Le, Tho Le-Ngoc, and Robert W Heath. Hybrid mmse precoding and combining designs for mmwave multiuser systems. *IEEE Access*, 5:19167–19181, 2017.

[101] Quentin H Spencer, A Lee Swindlehurst, and Martin Haardt. Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels. *IEEE transactions on signal processing*, 52(2):461–471, 2004.

[102] Max Costa. Writing on dirty paper (corresp.). *IEEE transactions on information theory*, 29(3):439–441, 1983.

[103] Martin Tomlinson. New automatic equaliser employing modulo arithmetic. *Electronics letters*, 7(5):138–139, 1971.

[104] Hiroshi Harashima and Hiroshi Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Transactions on Communications*, 20(4):774–780, 1972.

[105] Pulin Patel and Jack Holtzman. Analysis of a simple successive interference cancellation scheme in a ds/cdma system. *IEEE journal on selected areas in communications*, 12(5):796–807, 1994.

[106] Andrew LC Hui and K Ben Letaief. Successive interference cancellation for multiuser asynchronous ds/cdma detectors in multipath fading links. *IEEE Transactions on Communications*, 46(3):384–391, 1998.

[107] Ting-Ying Chang and Chiao-En Chen. A hybrid tomlinson–harashima transceiver design for multiuser mmwave mimo systems. *IEEE Wireless Communications Letters*, 7(1):118–121, 2018.

[108] Alberto Zanella, Marco Chiani, and Moe Z Win. Mmse reception and successive interference cancellation for mimo systems with high spectral efficiency. *IEEE Transactions on Wireless Communications*, 4(3):1244–1253, 2005.

[109] Peng Li, Rodrigo C De Lamare, and Rui Fa. Multiple feedback successive interference cancellation detection for multiuser mimo systems. *IEEE Transactions on Wireless Communications*, 10(8):2434–2439, 2011.

[110] Zhilong Zhang, Xiaoyong Wu, and Danpu Liu. Joint precoding and combining design for hybrid beamforming systems with subconnected structure. *IEEE Systems Journal*, 14(1):184–195, 2019.

[111] Jinho Choi. Analog beamforming for low-complexity multiuser detection in mm-wave systems. *IEEE Transactions on Vehicular Technology*, 65(8):6747–6752, 2015.

[112] Andrea Abrardo, Gábor Fodor, and Marco Moretti. Distributed digital and hybrid beamforming schemes with mmse-sic receivers for the mimo interference channel. *IEEE Transactions on Vehicular Technology*, 68(7):6790–6804, 2019.

[113] Ahmed Alkhateeb, Omar El Ayach, Geert Leus, and Robert W Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, 2014.

[114] Zhenyu Xiao, Tong He, Pengfei Xia, and Xiang-Gen Xia. Hierarchical codebook design for beamforming training in millimeter-wave communication. *IEEE Transactions on Wireless Communications*, 15(5):3380–3392, 2016.

[115] Junho Lee, Gye-Tae Gil, and Yong H Lee. Channel estimation via orthogonal matching pursuit for hybrid mimo systems in millimeter wave communications. *IEEE Transactions on Communications*, 64(6):2370–2386, 2016.

[116] Wei Wang and Wei Zhang. Orthogonal projection-based channel estimation for multi-panel millimeter wave mimo. *IEEE Transactions on Communications*, 68(4):2173–2187, 2020.

[117] Rakshith Rajashekar and Lajos Hanzo. Iterative matrix decomposition aided block diagonalization for mm-wave multiuser mimo systems. *IEEE Transactions on Wireless Communications*, 16(3):1372–1384, 2017.

[118] Chen Hu, Jian Liu, Xiangbai Liao, Yingzhuang Liu, and Jun Wang. A novel equivalent baseband channel of hybrid beamforming in massive multiuser mimo systems. *IEEE Communications Letters*, 22(4):764–767, 2018.

[119] Fuwang Dong, Wei Wang, and Zhenyu Wei. Low-complexity hybrid precoding for multi-user mmwave systems with low-resolution phase shifters. *IEEE Transactions on Vehicular Technology*, 68(10):9774–9784, 2019.

[120] Chanhong Kim, Taeyoung Kim, and Ji-Yun Seol. Multi-beam transmission diversity with hybrid beamforming for mimo-ofdm systems. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 61–65. IEEE, 2013.

[121] Hongyu Li, Ming Li, Qian Liu, and A Lee Swindlehurst. Dynamic hybrid beamforming with low-resolution pss for wideband mmwave mimo-ofdm systems. *IEEE Journal on Selected Areas in Communications*, 38(9):2168–2181, 2020.

[122] Feifei Gao, Bolei Wang, Chengwen Xing, Jianping An, and Geoffrey Ye Li. Wideband beamforming for hybrid massive mimo terahertz communications. *IEEE Journal on Selected Areas in Communications*, 39(6):1725–1740, 2021.

[123] Yiwei Sun, Zhen Gao, Hua Wang, Byonghyo Shim, Guan Gui, Guoqiang Mao, and Fumiyuki Adachi. Principal component analysis-based broadband hybrid precoding for millimeter-wave massive mimo systems. *IEEE Transactions on Wireless Communications*, 19(10):6331–6346, 2020.

[124] Didi Zhang, Yafeng Wang, Xuehua Li, and Wei Xiang. Hybrid beamforming for downlink multiuser millimetre wave mimo-ofdm systems. *IET Communications*, 13(11):1557–1564, 2019.

[125] José P González-Coma, Javier Rodriguez-Fernandez, Nuria González-Prelcic, Luis Castedo, and Robert W Heath. Channel estimation and hybrid precoding for frequency selective multiuser mmwave mimo systems. *IEEE Journal of Selected Topics in Signal Processing*, 12(2):353–367, 2018.

[126] Yun Chen, Da Chen, Tao Jiang, and Lajos Hanzo. Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave mimo systems. *IEEE Transactions on Communications*, 67(12):8315–8328, 2019.

[127] Cheng Wang, Edward KS Au, Ross D Murch, Wai Ho Mow, Roger S Cheng, and Vincent Lau. On the performance of the mimo zero-forcing receiver in the presence of channel estimation error. *IEEE transactions on wireless communications*, 6(3):805–810, 2007.