DESIGNING AND EVALUATING OF MIMO SYSTEMS FOR

MILLIMETER-WAVE COMMUNICATIONS

by
Mojtaba Ahmadi Almasi

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Boise State University

December 2020

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the **dissertation** submitted by

Mojtaba Ahmadi Almasi

Dissertation Title: Designing and Evaluating of MIMO Systems for Millimeter-Wave Communications

Date of Final Oral Examination: 11 December 2020

The following individuals read and discussed the dissertation submitted by student Mojtaba Ahmadi Almasi, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Hani Mehrpouyan, Ph.D.	Chair, Supervisory Committee
John Chiasson, Ph.D.	Member, Supervisory Committee
Hao Chen, Ph.D.	Member, Supervisory Committee

The final reading approval of the **dissertation** was granted by Hani Mehrpouyan, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

I dedicate my dissertation work to my lovely wife and parents.

ACKNOWLEDGMENTS

First and foremost, I wish to express my sincere appreciation to my advisor, Dr. Hani Mehrpouyan, who has the substance of a genius: he continuously supported and encouraged me to be professional and do the right thing even when the road got tough. I am thankful to him for giving me the freedom on choosing various, yet coherent research topics and directions and collaborators without objection. Dr. Mehrpouyan provided me a great opportunity to work with brilliant researchers during my Ph.D. work. I am also grateful for his valuable advice and great supervision that truly helped me improve my analysis, implementation, writing, and presentation skills which has made this project an unforgettable experience. Without his patience, motivation, enthusiasm, and immense knowledge, the goal of this project would not have been realized.

From the beginning days that I was working on my Ph.D. work, I have had this opportunity to closely work with Prof. Hamid Jafarkhani. I would like to thank Prof. Jafarkhani for his priceless advice and effort to provide me with his insightful comments that substantially improved the quality of my research papers and this dissertation. I also would like to gratitude him for hosting me at the Center for Pervasive Communications and Computing, University of California at Irvine. During my visit at University of California at Irvine, I had many technical meetings and fruitful discussions with Prof. Jafarkhani and I am thankful for such experience.

Bahareh. Without her unconditional support and encouragement, I could have never been able to finish this dissertation.

ABSTRACT

The fifth generation (5G) of wireless communications will integrate all existing technologies while bringing its own to the system. Amongst these technologies, millimeter-wave (mmWave) is emerging as a promising solution for 5G systems. However, to fully harness the potential of mmWave communications, obstacles such as severe path loss, channel sparsity, and hardware complexity should be overcome. The existing cost-effective systems can considerably reduce the hardware complexity and partially severe path loss, while channel sparsity still remains a main problem. Other factors such as transmission reliability and coverage area should be considered in 5G mmWave communications. Non-orthogonal multiple access (NOMA) is another enabling technology for 5G systems to improve spectral efficiency through serving more than one user at the same time/frequency/code resources. In particular, users' signals are superimposed in power domains at the transmitter which allows for simultaneously exploiting the available resources. Therefore, mmWave bands along with NOMA plays a crucial role in 5G wireless communications.

Aiming to overcome the mentioned obstacles, in the first part of this dissertation, we design a new lens-based reconfigurable antenna multiple-input multiple-output (RA-MIMO) architecture that takes advantage of multi-beam antennas for point-to-point communications. The considered antennas can generate multiple independent beams simultaneously using a single RF chain. This property, together with RA-

MIMO architecture, is used to combat small-scale fading and shadowing in mmWave bands. We use well-known space-time block codes (STBCs), together with phase-shifters at the receiver, in the RA-MIMO to suppress the effect of small-scale fading and shadowing. We also study the impact of practical quantized phase-shifters on the performance of the proposed RA-MIMO.

On the other hand, to make the most of these multi-beam antennas, a novel multiple access technique is developed for multi-user scenarios named reconfigurable antenna multiple access (RAMA). This technique transmits only each user's intended signal at the same time/frequency/code. This property makes RAMA an inter-user interference-free technique. Further, we integrate the well-known non-orthogonal multiple access (NOMA) technique in the proposed and other available mmWave systems. Moreover, to support a huge number of groups of users, we integrate RAMA into NOMA named reconfigurable antenna NOMA (RA-NOMA). This new technique divides the users with respect to their angle of departures (AoDs) and channel gains. Users with different AoDs and comparable channel gains are served via RAMA while users with the same AoDs but different channel gains are served via NOMA.

In the second part of this dissertation, we investigate NOMA in mmWave MIMO systems with phased array antennas. Two major obstacles in implementing NOMA are beam misalignment and limited channel coherence time due to the directional transmission. First, the effect of beam misalignment on rate performance in down-link of hybrid beamforming-based NOMA (HB-NOMA) systems is studied. To this end, an HB-NOMA framework is designed and a sum-rate maximization problem is formulated. An algorithm is introduced to design digital and analog precoders and efficient power allocation. Then, regarding perfectly aligned line-of-sight (LoS) chan-

nels, a lower bound for the achievable rate is derived. When the users experience misaligned LoS or non-LoS (NLoS) channels, the impact of beam misalignment is evaluated.

We take the limited channel coherence time into account for non-orthogonal multiple access (NOMA) in mmWave hybrid beamforming systems. Due to the limited coherence time, the beamwidth of the hybrid beamformer affects the beam-training time, which in turn directly impacts the data transmission rate. To investigate this trade-off, we utilize a combined beam-training algorithm. Then, we formulate a sum-rate expression which considers the channel coherence time and beam-training time as well as users' power and other system parameters.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1G	First Generation
4G	Fourth Generation
5G	Fifth Generation
AoA	Angle of Arrival
AoD	Angle of Departure
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base Station
CDMA	Code Division Multiple Access
CQI	Channel Quality Indicator
CRLH	Composite Right-Left Handed
CSI	Channel State Information
D-RAMA	Dynamic Power-Reconfigurable Antenna Multiple Access
EGC	Equal Gain Combining
EM	Electromagnetic
eMBB	Enhanced Mobile Broadband
FDMA	Frequency Division Multiple Access
F-RAMA	Fixed Power-Reconfigurable Antenna Multiple Access
HB-NOMA	Hybrid Beamforming Non-Orthogonal Multiple Access
ITU	International Telecommunication Union
LoS	Line-of-Sight
LWA	Leaky-Wave Antenna
MAC	Multiple Access Channel
MEMS	Microelectromechanical Systems
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error

mMTC	Massive Machine-Type Communication
mmWave	Millimeter Wave
MPC	Multi-Path Channel
MRC	Maximum Ratio Combining
NOMA	Non-Orthogonal Multiple Access
NLoS	Non-Line-of-Sight
OMA	Orthogonal Multiple Access
OSTBC	Orthogonal Space-Time Block Code
QAM	Quadratic Amplitude Modulation
QOSTBC	Quasi-Orthogonal Space-Time Block Code
RAMA	Reconfigurable Antenna Multiple Access
RA-MIMO	Reconfigurable Antenna Multiple Input Multiple Output
RF	Radio Frequency
SC	Superposition Coding
SER	Symbol Error Rate
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-Plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SPCA	Sequential Parametric Convex Optimization
STBC	Space-Time Block Code
TDMA	Time Division Multiple Access
TSA	Tapered Slot Antenna
ULA	Uniform Linear Array
URA	Uniform Planar Array
uRRLC	Ultra Reliable and Low Latency Communication
ZF	Zero Forcing

CHAPTER 1:

INTRODUCTION

Wireless communication is defined as a form of technology where information signals are sent and received through the air without using wire or cable. Within the last couple of decades, from the first generation (1G) to the forth generation (4G) of mobile communication, there have been rapid changes in technology, protocols, services offered or speed. The fundamental motivation behind this evolution is improving throughput, reliability, scalability, low latency, and energy-efficiency [3,4]. Due to the emergence of unprecedented services and applications such as autonomous vehicles, drone-based communication, wireless sensor networks [5], smart homes, and artificial intelligence-based personal assistants, the next generation of wireless communication should be compatible with the co-existence of human-centric and machine-type services and also a hybrid of the two. The fifth generation (5G) of wireless communication is expected to be a cure-all solution for such a diverse and complicated wireless environment. In order to guarantee that 5G will meet the expectations, the International Telecommunication Union (ITU) defined three representative service categories: (i) enhanced mobile broadband (eMBB), (ii) massive machine-type communication (mMTC), and (iii) ultra reliable and low latency communication (uRLLC) [6]. The key features of these service categories in the 5G systems have been summarized in Table 1.1 [2].

Table 1.1   **Summary of three service categories of 5G systems [2].**

Category	eMBB	mMTC	uRLLC
Main motivations	Increase spectral efficiency and increase peak throughput	Increase the number of support devices and support low-cost devices	Reduce end-to-end latency and increase robustness of data transmission
Key requirements	20x peak throughput 100x area throughput 5x spectral efficiency	10^6 devices/km^2 100x energy efficiency	10 msec end-to-end latency (e.g., 0.5 msec physical-layer latency)
Packet size	$\gg$ bytes	Hundreds of bytes	Tens to hundreds of bytes
Physical-layer solutions	Massive antennas, mmWave band-aggregation, and new waveforms	Non-orthogonal multiple access, overloaded data transmission, and active-user detection	Instant access, errorless data transmission
Applications	Virtual reality, real-time security, 3-dimensional image, and 4K-resolution video streaming	Logging, metering, lodging, tagging, and measuring	Tactile internet, remote control, self-car driving, argument reality, and industrial automation

1.1   MmWave Systems

One of the primary purpose of eMBB is to provide the end users higher data rates. Communication in mmWave bands is one of the promising approaches that can achieve a hundred to a thousand-fold increase in capacity over 4G systems. MmWave has a rich history in the wireless communication technology. The first study in mmWave bands at wavelengths as short as 5 to 6 mm was performed more than 100 years ago by Bose and Lebedew in the 1890s [7]. Further, the first application of the mmWave bands in mobile radio communication was invented in the 1990s and the 2000s, including system design and channel measurement [8–10]. The main reason to move toward mmWave communication is that the spectral resources in below 6GHz is becoming occupied while a plethora of unused spectrum is available in mmWave bands [11]. Despite the theoretical potential of mmWave communication, there are several key challenges for using mmWave in mobile networks, including severe path loss, shadowing, high penetration loss, limited channel coherence time, power-hungry devices, channel impairments, etc. In Chapters 3 and 4, we will address path loss,

shadowing, and high power consumption by designing a new system and algorithm using reconfigurable lens-based antennas. Further, in Chapter 6, we will study the limited channel coherence time.

1.2   NOMA

Non-orthogonal multiple access (NOMA) is another enabler for 5G systems. The fundamental premise behind NOMA in a single-cell network is to harvest all the benefits determined through information theory for the downlink transmission, modeled by the broadcast channel (BC), and uplink transmission, modeled by the multiple access channel (MAC). The capacity regions of the BC and MAC have been well-studied several decades ago, and concurrent non-orthogonal transmission is the optimal transmission strategy in both channels [12]. In order to increase the spectral efficiency and support a large group of users, NOMA is introduced as a promising solution to mmWave 5G systems. Despite its capability in mmWave bands, exploiting NOMA requires overcoming several challenges such as high path loss, beam misalignment, limited channel coherence time, hardware complexity at the user end, high power consumption, and etc. To this end, in Chapter 4, we will propose new multiple access techniques to reduce hardware complexity and increase spectral efficiency. In Chapter 5, we will study the impact of beam misalignment. In Chapter 6, we concentrate on the limited channel coherence time issue.

1.3   Summary of Contributions

In Chapter 3, we propose a novel MIMO architecture exploiting lens-based reconfigurable antennas to overcome small-scale and large-scale fading and keep the hardware complexity and costs low. The main contributions of this chapter are: 1) A new lens-

based MIMO architecture is proposed for point-to-point mmWave communications. Instead of using a large number of phase-shifters for beamforming, we use a lens antenna along with a switching network. Further, a channel model that describes the relationship between the transmit and receiver signal is derived. 2) To improve reliability of mmWave links, we use rate-one STBCs. Our motivation to use STBCs is their advantage in reducing the error probability without requiring full channel state information (CSI) at the transmitter. The obtained diversity gain can mitigate small-scale fading and shadowing effects. 3) We evaluate the diversity gain and symbol error rate (SER) performance of the proposed system for ideal and digital phase-shifters assuming only small-scale fading. Further, an upper bound on SER is derived for the ideal and digital phase-shifters. 4) We evaluate outage probability when the impact of shadowing is included. The analysis reveals that RA-MIMO prevents deep shadowing through using several channels for the transmission. The results of this work are published in [13–15].

In Chapter 4, we propose a fresh multiple access technique for mmWave reconfigurable antenna systems which is called reconfigurable antenna multiple access (RAMA). RAMA aims to send only the desired signal of each user whereas simultaneously communicating with many users. RAMA achieves a higher sum-rate than both OMA and NOMA. Since the operation of RAMA is limited to only one user per beam, we propose a new multiuser technique which integrates RAMA and NOMA, and is referred to as RA-NOMA. In RA-NOMA, each antenna element is connected to all RF chains. Users are grouped into two different sets based on the angle of departures (AoDs) and the users' channel gain. The users with different AoDs and comparable channel gains are supported via RAMA while the users with the same

AoDs but different channel gains are served via NOMA. This technique introduces a new degree of freedom for grouping the users without adding RF chains. The major contribution of this chapter are: 1) We consider a scenario in which a single BS is equipped with a mmWave reconfigurable antenna and each beam of the antenna serves one user where the users are not aligned with the same direction. Given that the limitation on the RF circuitry of the antenna results in the division of the transmitted power amongst the beams, the current state-of-the-art in mmWave-NOMA would not operate efficiently in such a setting. To enhance the performance of multiple access schemes in the mmWave band and also overcome this fundamental limit for reconfigurable antennas, RAMA technique is proposed. Since RAMA aims to transmit only the intended signal of each user, SC and SIC are not required. 2) Although RAMA achieves a higher sum-rate than NOMA, its operation is limited to only one user per beam. To overcome this limit, we propose a new multiuser technique which integrates RAMA and NOMA, and is referred to as RA-NOMA. In RA-NOMA, each antenna element is connected to all RF chains. Users are grouped into two different sets based on the AoDs and the users' channel gain. The users with different AoDs and comparable channel gains are supported via RAMA while the users with the same AoDs but different channel gains are served via NOMA. This technique introduces a new degree of freedom for grouping the users without adding RF chains. The results of this work are published in [14, 16, 17]

In Chapter 5, we study the effect of integration of hybrid beamforming and NOMA on the achievable rate in the presence of beam misalignment. The contributions of this chapter are: 1) We incorporate the 5G enabling technology NOMA and a multiuser HB system studied in [18]. Since we aim to evaluate the impact of beam misalign-

ment on the downlink of HB-NOMA systems, a sum-rate expression is formulated. Specifically, we revise the sum-rate expression in [18] with regard to the NOMA technique. Then, an algorithm is introduced to maximize the system sum-rate subject to a total power constraint, in three steps. To get the first and second steps, we design the analog and digital precoders only regarding LoS channels using the well-known strong effective channel-based effective channel precoder. The third step is a location-based static power allocation. 2) As the maximized sum-rate directly depends on the effective channels of users, we first study the rate for perfect beam alignment where all users exploit LoS channels. A lower bound is derived for the achievable rate of an HB-NOMA user. The bound reveals that the interference is just due to using NOMA in which the SC technique at the transmitter and SIC at the receiver are exploited. That is to say, the interference on a user is caused by NOMA users located inside the same cluster, which is called an intra-cluster interference. Indeed, HB slightly amplifies the noise term which is led by analog devices used in the beamformer. The analysis shows that for the perfect alignment, the HB-NOMA users can achieve a rate which is close to that of NOMA with the fully-digital beamforming systems. 3) We study the achievable rate of the maximized sum-rate for misaligned beams between the BS and users in the presence of misaligned LoS and NLoS channels. Toward this goal, the beam misalignment problem is modeled by a beam misalignment factor. Considering the derived factor, the effective channel of the users with misaligned LoS or NLoS channel is described in terms of the aligned effective channel parameter and the misalignment factor. 4) We extract a lower bound for the achievable rate using the effective channel model. Three terms, i.e., intra-cluster interference, inter-cluster interference, and noise, constrain the achievable rate. Unfortunately, these terms are

directly or indirectly associated with misalignment factors. It is concluded that in HB-NOMA with the precoder based on the strongest effective channel the achievable rate of a user depends on both the effective channel gain and beam alignment issue. This is opposite to the fully-digital NOMA systems in which only the effective channel gain affects the rate. Then, an upper bound for rate gap between the aligned and misaligned HB-NOMA user is found. The results of this work are published in [19,20].

In Chapter 6, we jointly optimize the beamwidth and power in hybrid beamforming-based mmWave-NOMA systems. In practical scenarios, neglecting the effect of beam-training duration may cast doubt on the performance of the mmWave-NOMA networks. Especially, since the channel coherence time in mmWave bands is limited [21], the beam-training duration should be adequately small. Thus, on one hand, a small beam-training duration results in a wide beamwidth, i.e., low beamforming gain, and noisy channel estimation. On the other hand, a long beam-training time provides robust beamforming and accurate channel estimation but imposes a delay in data transmission. This may not be desirable in delay-sensitive systems as it leaves less time for data transmission and leads to low sum-rates. The contributions of this chapter are: 1) We consider the well-studied mmWave hybrid beamforming combined with NOMA for limited coherence time scenarios. The system can control the beamwidth, using the phase-shifters deployed in the hybrid beamformer, and allocate power to NOMA users. To this end, a tone-based beam-training algorithm [22] compatible to our mmWave-NOMA system is utilized. The algorithm combines the exhaustive search [23] and tone-based beam-training [22] algorithms. 2) Unlike the existing multi-beam mmWave-NOMA systems, we take the channel coherence time into account. The limited coherence time leads to a trade-off between the beamwidth

resolution and the data transmission rate. We also formulate a new sum-rate expression for optimization. 3) A joint power and beamwidth optimization algorithm is proposed which iterates between the power allocation and beamwidth optimization. The results of this work are presented in [24].

1.4 List of Publications

The above contributions, so far, has led to the following publications in prestigious journals and conferences:

1.4.1 Journal Papers

- **M. Ahmadi Almasi**, L. Jiang, H. Jafarkhani, and H. Mehrpouyan, "Joint Beamwidth and Power Optimization in MmWave Hybrid Beamforming-NOMA Systems," *Accepted in IEEE Trans. Wireless Commun.*

- **M. Ahmadi Almasi**, R. Amiri, H. Mehrpouyan, and H. Jafarkhani, "MmWave Lens based MIMO System for Suppressing Small-scale Fading and Shadowing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5292-5306, Aug. 2020.

- **M. Ahmadi Almasi**, M. Vaezi, and H. Mehrpouyan, "Impact of Beam Misalignment on Hybrid Beamforming NOMA for mmWave Communications," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4505-4518, June 2019.

1.4.2 Conference Papers

- **M. Ahmadi Almasi**, R. Amiri, M. Vaezi and H. Mehrpouyan, "Lens-based mmWave Reconfigurable Antenna NOMA," *in Proc. IEEE Int. Conf. Commun. (ICC) Workshops*, pp. 1-5, May 2019.

- **M. Ahmadi Almasi** and H. Mehrpouyan, "Non-orthogonal multiple access

based on hybrid beamforming for mmwave systems," *in Proc. IEEE Veh. Technol. Conf. (VTC),* Fall, Aug. 2018.

- **M. Ahmadi Almasi**, H. Mehrpouyan, V. Vakilian, N. Behdad, and H. Jafarkhani, "A New Reconfigurable Antenna MIMO Architecture for mmWave Communication," *in Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1 – 7, May 2018. (**Best paper award**)

- **M. Ahmadi Almasi**, H. Mehrpouyan, D. Matolak, C. Pan, and M. Elkashlan, "Reconfigurable Antenna Multiple Access for 5G mmWave Systems," *in Proc. IEEE Int. Conf. Commun. (ICC) Workshops* , pp. 1 – 6, May 2018.

1.4.3 Patent

M. Ahmadi Almasi and H. Mehrpouyan, "Reconfigurable Antenna Multiple Access for Millimeter Wave Systems," US16209350.

1.5   List of Awards

- The exemplary Reviewer of IEEE Commun. Lett. in 2019.

- Best Paper Award, IEEE Int. Conf. Commun. (ICC), May 2018.

- IEEE ComSoc Travel Grant to attend Int. Conf. Commun. (ICC), May 2018.

- Graduate Student Showcase Winner, Boise State University, April 2018.

1.6   Dissertation Organization

This dissertation is organized as follows. In Chapter 3, a novel mmWave lens-based MIMO architecture is proposed. Chapter 4 proposes two new multiple access techniques called RAMA and RA-NOMA. In Chapter 5, the impact of beam misalignment

on the performance of hybrid beamforming-based NOMA is investigated. Chapter 6 jointly designs beamwidth and power allocation for HB-NOMA systems considering limited channel coherence time.

Notation: For the rest of this dissertation, the models and problems in each chapter has been developed, analyzed, and solved comprehensively and independently. Thus, the notations used in each chapter is specific for that chapter and is not extended to the other chapters.

CHAPTER 2:

BACKGROUND

In this chapter, the concepts and ideas that are common for the dissertation will be covered. We first discuss the millimeter-Wave (mmWave) beamforming systems. Then, the mmWave channel model is explained. Finally, non-orthogonal multiple access (NOMA) is described.

2.1 MmWave Beamforming Designs

MmWave technology operating in the 30-300 GHz range is emerging as a promising solution for the fifth generation (5G) of wireless communication systems by supporting a larger user base and higher speed links [25]. The existence of a large communication bandwidth at mmWave frequencies will enable mmWave systems to support multi Gigabits/sec speeds. However, significant large-scale fading, i.e., path loss and shadowing, channel sparsity in multiple input multiple output (MIMO) systems, small-scale fading, as well as hardware limitations and costs are all major obstacles for the deployment of mmWave systems.

To combat the severe path loss at mmWave frequencies, researchers have proposed using antennas with substantial directional gains and line-of-sight (LoS) links [26]. Directivity can be achieved by using large number of antennas at transmitters and receivers. Therefore, thanks to the short wavelength at mmWave frequencies, massive

MIMO is used to mitigate severe path loss [27–42]. Regarding this, three beamforming architectures have been proposed for mmWave systems with massive MIMO: i) *digital* [27–29], ii) *analog* [30–32], and iii) *hybrid* [33–43].

The *digital beamforming* approach provides great flexibility in shaping the transmitted beams. However, it requires one radio frequency (RF) chain per antenna [27–29]. This results in significant cost and complexity in mmWave massive MIMO systems. Further, digital beamforming may result in significant delay due to a large number of channel parameters that must be estimated [27]. As an alternative, *analog beamforming* applies phase-shifters to shape the output beam with only one RF chain for all antennas [30–32]. Although analog beamforming is energy efficient and cost effective, it can only provide one highly directional beam which does not address the high data rate required for outdoor 5G networks [27]. A promising beamforming approach for the mmWave MIMO architecture is based on a combination of analog and digital beamforming, i.e., *hybrid beamforming*. The hybrid architecture aims to use the advantages of digital and analog architectures. That is, at low signal-to-noise ratio (SNR), the hybrid architecture performs only beamforming to combat the path loss, and at high SNR, it aims to transfer high data by performing multiplexing [31]. Due to significance of hybrid beamforming, in what follows, we review the existing architectures.

To the best of the authors' knowledge, three types of hybrid beamforming architectures have been proposed so far which are i) hybrid beamforming with phase-shifters [33–36], ii) hybrid beamforming with spherical and planar lens antennas [37–41,43], and iii) hybrid beamforming with reconfigurable[1] microelectromechanical sys-

---

[1]In general, the reconfigurable antennas are designed to reconfigure their operational frequency [44,45], polarization [46], and/or pattern (beam) [47].
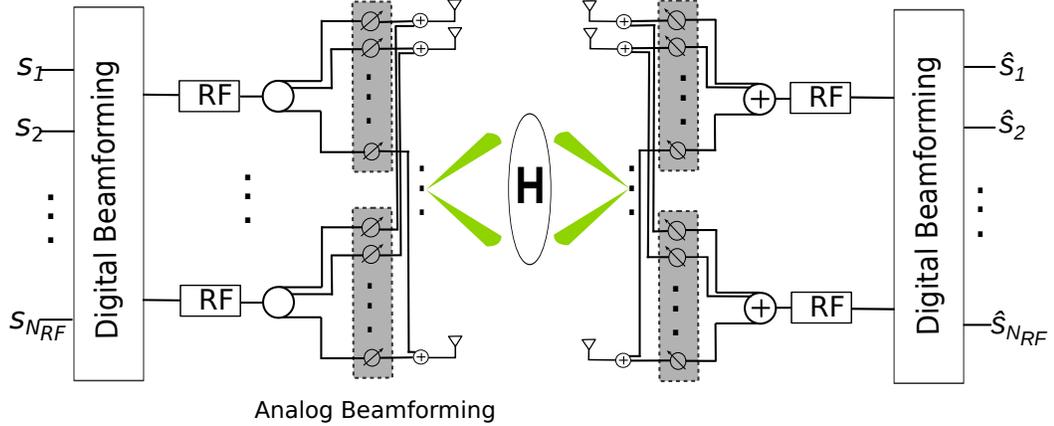
Figure 2.1 Architecture of hybrid beamforming.

tems (MEMS) integrated antennas [42]. In the first architecture shown by Fig. 2.1, each RF chain is connected to all the antennas via phase-shifters to generate highly directional beams [33, 34]. The system aims to achieve multiplexing gain in order to simultaneously transmit multiple information streams. Hence, the throughput is increased. However, it is indeed complex due to huge number of phase-shifters. To reduce the complexity, the concept of sub-connected network has been introduced in [35]. In this architecture, each RF chain is connected to a sub-group of antennas, which reduces the complexity at the cost of reducing the directionality of the beams. Later, [36] proposed a phase-shifter selection design to further reduce the complexity.

The second methodology has a completely different structure, where several RF chains are connected to a *lens antenna array* via a switching network known as *beamspace* MIMO [37]. In this architecture shown by Fig. 2.2, instead of phase-shifters, a spherical lens antenna is used to constructively change the phase of each path and generate directional beams. Hence, without using phase-shifters, the transceiver antenna can generate a few orthogonal beams to achieve multiplexing gain for better utilization of the bandwidth and to preserve low hardware complexity. The multi-
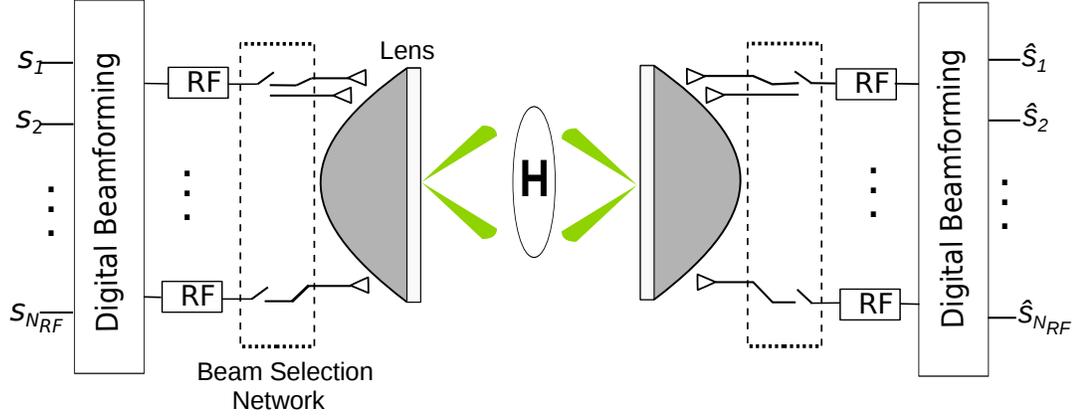
Figure 2.2   Architecture of beamspace MIMO system.

plexing gain highly depends on the the aperture size and the distance between the transmitter and receiver. Such that, for a fixed aperture size, by increasing the distance, the LoS channel matrix might be ill-conditioned and the multiplexing gain is degraded. In addition, several works have studied the lens antenna array in mmWave systems [38–41]. In [38], an electromagnetic (EM) planar lens is integrated with an antenna array. In the uplink transmission, the EM lens focuses the energy of the desired signal and rejects the interference. This leads to performance gain and lower signal processing complexity. Zeng and Zhang [39] then proposed a new lens antenna array MIMO system and demonstrated that the array response follows a sinc function. Next, an RF lens-embedded massive MIMO system is investigated in [40]. The fabricated lens operates at the 77-GHz band and is shown to support a high gain in the desired direction. The work in [41] studied a cost-effective mmWave lens antenna array system. The authors show that by exploiting the angle-dependent energy, together with the angular sparsity of the mmWave channels, the wireless system can achieve optimal capacity, i.e., the capacity achieved by digital beamforming, with only a few RF chains and low-complexity single-carrier transmission.

The third methodology offers reconfigurable MEMS integrated antennas for mmWave frequencies. The MEMS-based reconfigurable antennas have been widely studied in sub-6 GHz systems [48–51]. In [49], the diversity gain of traditional MIMO systems are improved through the application of reconfigurable antennas at the receiver and space-time block codes (STBCs). Subsequently, [50] extends the technique proposed in [49] to a system with reconfigurable antennas at both transmitter and receiver sides. Later, a coding scheme was proposed for these MIMO systems over frequency-selective fading channels in [51]. Due to high hardware complexity, costs, and propagation properties of mmWave channels, the designed systems in [49–51] are not viable for mmWave systems. Regarding these antennas, recently, a hybrid MEMS-based reconfigurable antenna system has been suggested for mmWave transmission in [42]. It is shown that the MEMS-based reconfigurable antennas achieve high throughput gains.

2.2 MmWave Propagation Channel Model

We use the widely adopted extended Saleh-Valenzuela model as a multi-path channel (MPC) model throughout this dissertation [33,52]. In this model, each LoS and NLoS path is defined by a channel gain and an array steering vector at the transmitter and an array response vector at the receiver. Hence, the channel matrix between the transmitter and the receiver for a uniform planar array (UPA) is given by

$$\mathbf{H} = \frac{1}{\sqrt{L+1}} \left(\beta_0 \mathbf{G}_0 + \sum_{l=1}^{L} \beta_l \mathbf{G}_l \right), \tag{2.1}$$

where β_0 and β_l denote the channel fading coefficient of LoS and NLoS channels, respectively. $\mathbf{G}_0 \in \mathbb{C}^{N_r \times N_t}$ is the LoS channel matrix and $\mathbf{G}_l$ is the lth NLoS channel

matrix. Further, N_t and N_r denote the number of antennas at the transmitter and receiver, respectively. In particular, $\mathbf{G}_l$, $0 \leq l \leq L$, is given by

$$\mathbf{G}_l = \mathbf{a}_\text{r}(\theta_l^\text{az}, \theta_l^\text{el})\mathbf{a}_\text{t}^\dagger(\phi_l^\text{az}, \phi_l^\text{el}), \tag{2.2}$$

where θ_l^az (θ_l^el) and ϕ_l^az (ϕ_l^el) are normalized azimuth (elevation) AoA and angle of departure (AoD), respectively. Also, $\mathbf{a}_\text{t} \in \mathbb{C}^{N_\text{t} \times 1}$ and $\mathbf{a}_\text{r} \in \mathbb{C}^{N_\text{r} \times 1}$ are the antenna array steering vector and array response vector of the transmitter and the receiver, respectively. Suppose the UPA is of dimension $N_x \times N_y$ in which N_x (N_y) denotes the number of elements of x-axis (y-axis) at the transmitter. The array steering vector $\mathbf{a}_\text{t}(\phi_l^\text{az}, \phi_l^\text{el})$ can be defined as

$$\mathbf{a}_\text{t}(\phi_l^\text{az}, \phi_l^\text{el}) = \frac{1}{\sqrt{N_x N_y}} \left[1, \ldots, e^{-j\pi\phi_{m,n}}, \ldots, e^{-j\pi\phi_{N_x-1, N_y-1}} \right]^\dagger, \tag{2.3}$$

where $\phi_{m,n} = \frac{2d'}{\lambda} \left(m\sin\phi_l^\text{az}\cos\phi_l^\text{el} + n\sin\phi_l^\text{az}\sin\phi_l^\text{el} \right)$ for $m \in \{0, \ldots, N_x - 1\}$ and $n \in \{0, \ldots, N_y - 1\}$. λ and d' are the wavelength and antenna spacing, respectively. Each path or cluster consists of $N_\text{ray} = N_x \times N_y$ rays. The array response vector can be obtained at the same fashion.

2.3   NOMA

NOMA is another enabling technique for 5G networks that augments the number of users and spectral efficiency in multiuser scenarios [53–60]. Unlike orthogonal multiple access (OMA) techniques, such as time division multiple access (TDMA), frequency division multiple access (FDMA), and code division multiple access (CDMA) which can support only one user per time, frequency, or code, respectively, NOMA can sup-

port multiple users in the same time/frequency/code/beam. NOMA can be realized in the code, power, or other domains [12]. In the power domain, NOMA employs *superposition coding* at the transmitter. This technique exploits the channel gain difference between users to multiplex their signal. Subsequently, *successive interference cancellation* (SIC) is applied at the receiver such that the user with the better channel first decodes the signal of the user with worse channel and then subtracts it from the received signal to decode its own signal [12,53–61]. Besides superiority of NOMA over OMA techniques in terms of number of supported users and spectral efficiency, OMA techniques may not be a practical option for mmWave communications [62]. As an example, TDMA, which serves users through orthogonal time slots but the same spectrum, requires precise and fast timing synchronization. This is because symbol rate in 5G networks is far higher than the current networks. Therefore, employing TDMA in a mmWave 5G network might be challenging. Exploiting FDMA in mmWave 5G network also can bring about implementation issues. In FDMA, the existing large frequency band is divided into several orthogonal frequency bands. It is expected that FDMA serve all users via the orthogonal bands at the same time slot. However, due to highly directional beams, the current mmWave systems are not able to cover all users' locations and only a few users will be supported. Further, frequency band division causes the allocated bandwidth for each user in a dense mmWave network to become small. So, mmWave networks may not have enough bandwidth to support the users with the required high data-rate. The obstacles related to using CDMA in mmWave frequencies have been explained in [62]. The propagation characteristics of mmWave frequencies are another reason to incorporate the hybrid beamforming systems and NOMA. Transmission in mmWave band suffers from high path loss and

thus users in different locations may experience very different channel gains. This implies that the mmWave band better suits power domain NOMA which offers a larger spectral efficiency when the channel gain difference between the users is high. Severe shadowing and blockage are other factors that make mmWave links vulnerable to outage [25, 26, 63]. Although the large unused spectrum in mmWave bands is envisioned as a promising solution for high data-rate transmission in 5G networks, high path loss and outage due to shadowing and blockage make mmWave links prone to temporary shutdowns. Hence, when the link exists, increasing the spectral efficiency will lead to higher data-rates. This would meet the required unprecedented throughput of $1000\times$ current networks in 5G networks.

To describe the general concept of NOMA, it is assumed that there are only two users and an omni-directional antenna is employed at the base station (BS) and users. The users are defined by User 1 and User 2. Assume the BS sends the signals s_k for User k, $k = 1, 2$, where $\mathbb{E}[|s_k|^2] = 1$, with power transmission p_k. The sum of p_ks equals to p. According to the principle of NOMA downlink, at the transmitter, s_1 and s_2 are superposition coded as

$$x = \sqrt{p_1}s_1 + \sqrt{p_2}s_2. \tag{2.4}$$

Hence, the received signal at User i is given by

$$y_k = h_k x + n_k \equiv h_k(\sqrt{p_1}s_1 + \sqrt{p_2}s_2) + n_k, \tag{2.5}$$

where h_k is the channel gain containing the small scale fading and path loss. Also, n_k denotes the additive white Gaussian noise with power σ^2. At the receiver, each user

performs the SIC process to decode the desired signal. The optimal decoding order depends on the channel gain. Without loss of generality, let us assume that User 1 is closer to the BS, i.e., $d_1 \leq d_2$, which gives $p_1 \leq p_2$ [64].

After applying SIC, User 1 removes the term $\sqrt{p_2}s_2h_1$. Hence, the achievable rate of the NOMA users can be determined as, respectively,

$$
\begin{cases}
R_1 = \log_2(1 + \rho_1|h_1|^2), \\
R_2 = \log_2(1 + \frac{\rho_2|h_2|^2}{\rho_1|h_2|^2+1}),
\end{cases}
\tag{2.6}
$$

where $\rho_k = \frac{p_k}{\sigma^2}$ is defined as signal-to-noise ratio (SNR) of the User k. This result indicates that in NOMA power allocation plays a crucial rule. For example, an improper power allocation does not allow User 1 to cancel the term $\sqrt{p_2}s_2h_1$ completely, which in turn does not allow the interference from User 2 to be successfully eliminated.

CHAPTER 3:

MMWAVE LENS-BASED MIMO SYSTEM FOR SUPPRESSING SMALL-SCALE FADING AND SHADOWING

In this chapter, we propose a generalized millimeter-Wave (mmWave) reconfigurable antenna multiple-input multiple-output (RA-MIMO) architecture that takes advantage of lens antennas. The considered antennas can generate multiple independent beams simultaneously using a single RF chain. This property, together with the RA-MIMO, is used to combat small-scale fading and shadowing in mmWave bands. To this end, first, we derive a channel matrix for RA-MIMO. Then, we use rate-one space-time block codes (STBCs), together with phase-shifters at the receive reconfigurable antennas, to suppress the effect of small-scale fading. We consider two kinds of phase shifters: i) ideal which is error-free and ii) digital which adds quantization error. The goal of phase-shifters is to convert a complex-valued channel matrix into real-valued. Hence, it is possible to use rate-one STBCs for any dimension of RA-MIMO. We investigate diversity gain and derive an upper bound for symbol error rate in cases of ideal and digital phase-shifters. We show that RA-MIMO achieves the full-diversity gain with ideal phase-shifters and the full-diversity gain for digital phase-shifters when the number of quantization bits is higher than one. We investi-

gate RA-MIMO in the presence of shadowing. Our analysis demonstrates that, by increasing the dimension of RA-MIMO, the outage probability decreases which means the effect of shadowing decreases.

3.1 Introduction

All hybrid architectures described in Section 2.1 aim to reduce the hardware complexity and costs, partially resolve the path loss issue through beamforming gain, and increase the spectral efficiency through multiplexing gain. Nevertheless, there are remaining essential issues related to large-scale fading, small-scale fading, and channel sparsity that must be addressed [25, 26]. The current mmWave systems only concentrate on transferring high data rate at high SNRs through multiplexing gain, i.e., steering a few directional beams, and mitigating path loss at low SNRs through beamforming, i.e., steering a highly directional and strong beam [27–42]. Thus, the current mmWave transmission relies on sending each data stream through one channel. In what follows, we address the deficiencies of the current mmWave transmission systems.

Point-to-Point transmission in mmWave dense networks is usually power-limited (or equally noise-limited) [63, 65]. Moreover, due to high path loss and blockage, LoS channels are available at a short distance and are susceptible to outage probability [63, 65]. On the other hand, because of small-scale fading and large-scale fading, NLoS channels suffer from signal attenuation and fast fluctuations [25, 66, 67]. As such, experimental channel measurements reveal that mmWave NLoS channels experience log-normal shadowing with zero mean and variances of 8.7 and 8 dB at 28 and 73 GHz [67], respectively. Thus, even at high SNRs, transmission via one NLoS channel is unreliable and susceptible to deep fading which leads to frequent outages. Notice

that in this chapter, shadowing is defined as large-scale fluctuations in the received signal which is due to the scatterers. Shadowing due to buildings, human body and mobile objects which is common in LoS channels is defined as blockage. The blockage can be severe and may result in the received signal power to be below the noise floor level [67,68]. It is worth mentioning that one technique to mitigate the effect of shadowing is micro-diversity [69,70]. In this technique, whenever the channel between a user and a base station (BS) is in deep shadowing, the user is served by another BS with a better channel to the user. This technique is utilized in both sub-6 GHz [71] and mmWave frequencies [68–70]. In mmWave, the fast-changing shadowing situation will require frequent hand-over between the BSs which imposes excessive overhead. More importantly, LoS channels may not be always available between the user and the neighboring BSs. For instance, in [70] it is shown that when a user is connected to one BS, in a network with the density of 10^3 BSs per km^2, the probability of having one LoS channel is 0.9 (< 1).

3.1.1   Contributions

To tackle small-scale and shadowing issues in mmWave bands, one viable solution is achieving channel diversity gain which is the subject of this chapter. Toward this goal, lens-based reconfigurable antennas are deployed to design a new mmWave MIMO system. Recall that traditional sub-6 GHz MIMO systems can exploit beam-forming, multiplexing gain, and diversity gain [71]. The diversity gain is a well-known technique for overcoming small-scale fading. To resolve shadowing, we notice that, fortunately, each cluster (channel) in mmWave systems experiences random and independent shadowing [66–68, 72]. The randomness and independence of clusters can make the proposed RA-MIMO suitable to overcome the impact of shadowing. To

this end, recently, the idea of using a single composite right-left handed (CRLH) leaky-wave antenna (LWA) in a MIMO system instead of a massive antenna array was proposed in [73]. In this work, which is limited to just 2×2 MIMO systems, two reconfigurable LWAs are deployed at the transmitter in which each antenna steers several beams toward the receiver that is equipped with omni-directional antennas. Nevertheless, the recent designed lens antennas in [74] and [1] might represent better characteristics compared to that of LWAs that was considered in [73]. In this chapter, inspired by [74] and [1], we propose an $N_t \times N_r$ lens-based reconfigurable antenna MIMO (RA-MIMO) architecture for mmWave communications. It is worth mentioning that mainly there are three differences between the proposed MIMO and hybrid beamforming with MEMS-based reconfigurable antennas. First, each lens-based reconfigurable antenna can steer multiple independent beams, while this unique property does not hold for MEMS-based reconfigurable antennas [1, 16, 17]. That is, the number of required RF chains in MEMS-based reconfigurable antennas is the same or greater than the number of data streams [42]. Second, in hybrid beamforming with MEMS-based reconfigurable antennas, all RF chain are connected to an antenna array [42], whereas in the proposed MIMO RF chains are separate and each RF chain is connected to a lens-based antenna. Third, the MEMS-based reconfigurable antennas can change the antenna's radiation pattern, polarization, and frequency [42, 48, 75]. However, here, it is assumed that the lens-based antennas only reconfigure their radiation pattern. Although the proposed MIMO architecture can reconfigure radiation polarization and frequency by adequately designing the tapered slot antennas (TSAs). Further, the proposed system aims to overcome small-scale and large-scale fading and keep the hardware complexity and costs low. The main contributions of this chapter

are:

1. A new lens-based MIMO architecture is proposed for point-to-point mmWave communications. Instead of using a large number of phase-shifters for beam-forming, we use a lens antenna along with a switching network. Further, a channel model that describes the relationship between the transmit and receiver signal is derived.

2. To improve reliability of mmWave links, we use rate-one STBCs. Our motivation to use STBCs is their advantage in reducing the error probability without requiring full channel state information (CSI) at the transmitter. The obtained diversity gain can mitigate small-scale fading and shadowing effects.

3. We evaluate the diversity gain and symbol error rate (SER) performance of the proposed system for ideal and digital phase-shifters assuming only small-scale fading. Further, an upper bound on SER is derived for the ideal and digital phase-shifters.

4. We evaluate outage probability when the impact of shadowing is included. The analysis reveals that RA-MIMO prevents deep shadowing through using several channels for the transmission.

5. Numerical simulations show that the proposed architecture outperforms hybrid beamforming systems in terms of diversity gain. Indeed, RA-MIMO with digital phase-shifters mostly achieves full-diversity at the cost of a slight coding gain loss. The impact of the number of transmit antennas on the received SNR and BER is presented.

The rest of the chapter is organized as follows: Section 3.2 presents the channel model, the proposed RA-MIMO structure, and system parameters. Section 3.3 addresses small-scale fading by designing STBCs, evaluates full-diversity and error rate performance, and represents the system design tradeoffs. Section 3.4 evaluates outage probability in presence of shadowing. In Section 3.5, we present simulation results. Finally, Section VI concludes the chapter.

Notations: Hereafter, $j = \sqrt{-1}$, small letters, bold letters and bold capital letters will designate scalars, vectors, and matrices, respectively. Superscript $(\cdot)^{\dagger}$ and $(\cdot)^{*}$ denote the transpose and transpose-conjugate operators, respectively. The sign $\angle(\cdot)$ stands for the phase of $(\cdot)$. $(\cdot)^{\mathrm{Re}}$ and $(\cdot)^{\mathrm{Im}}$ denote the real and imaginary parts of $(\cdot)$, respectively. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Next, $\mathrm{vec}(\cdot)$ denotes the vectorization of $\mathbf{A}$ which is a column vector obtained by stacking the columns of the matrix on top of one another. Further, $\mathbb{E}(\cdot)$, $|\cdot|$, and $||\cdot||^2$ denote the expected value, absolute value, and norm-2 of $(\cdot)$, respectively. $(\cdot)^{-1}$ and $\mathrm{tr}(\cdot)$ denote the inverse and the trace of matrix $(\cdot)$, respectively. $Q(x)$ denotes the Q-function and is defined as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{u^2}{2}\right) du$. Finally, $\mathbf{A} \circ \mathbf{B}$ and $\mathbf{A} \otimes \mathbf{B}$ stand for the Hadamard product and the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively.

3.2   The Proposed RA-MIMO Structure

3.2.1   Channel Model

MmWave propagation channel model

We consider the channel described in (2.1) with D channels. It is worth mentioning that D represents the number of propagation paths established via NLoS channels which are due to scatterers between the transceiver. In practice, mmWave LoS chan-

nels exist in short distances and variation of the received signal due to small-scale fading and shadowing is negligible. In contrast to LoS channels, mmWave NLoS channels are impaired by small-scale and large-scale fading. Further, it is worth noting that the probability of having NLoS channels in mmWave frequencies, e.g., at 28 and 73 GHz, is much higher than LoS channels [67].

Channel model between a transceiver lens antenna pair

In what follows, we derive the channel model between two lens-based reconfigurable antennas. Before this, let us introduce a lens-based reconfigurable antenna. Each antenna contains four main functional blocks: i) **RF transceiver chain** which modulates and demodulates data signal. ii) **Beam selection network** is connected to the output of the RF chain and has multiple outputs that are connected to the input ports of the TSA feeds. iii) **TSA feeds** are located on the focal surface of a spherical lens. Each TSA feed steers a beam. iv) **A spherical lens** is positioned in front of the TSA feeds. Note that, for simplicity, hereafter, we use the terms TSA, lens, and lens antenna instead of TSA feeds, a spherical lens, and lens-based reconfigurable antenna, respectively. Further, N_{TSA} TSAs are located on the surface of each lens. The lens acts as a passive phase-shifter network. Each plane wave goes through the lens. After shifting the plane wave with different delays by the lens, the energy of the plane wave is focused on one of the focal points. When a plane wave hits the surface of a lens, the incident plane wave is focused on the antipodal point on the lens [76,77]. Accordingly, when a feed antenna on the focal point sends a wave through the lens, a plane wave departs the lens as in Fig. 3.1. We now define each plane wave with the channel gain, direction information, and N_{ray} rays. Therefore, the model $\mathbf{H}_0$ is a
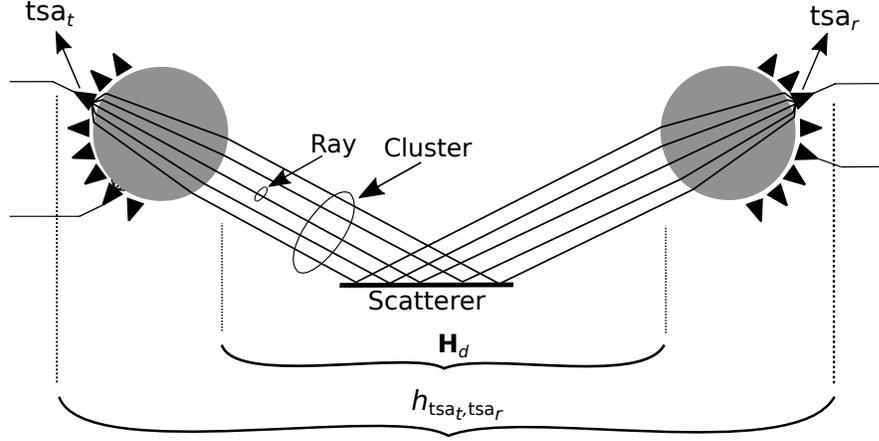
Figure 3.1: Two different types of channel for a single path. The channel matrix $\mathbf{H}_d$ describes the dth channel between transmit lens antenna and receive lens antenna based on the cluster-based Saleh-Valenzuela model. h_{tsa_t,tsa_r} denotes the channel between TSA tsa_t located on the transmit lens and TSA tsa_r on the receive lens.

suitable model to describe the channel between two lenses.

Channel model between a transceiver TSA pair

The channel impulse response for the dth path, i.e., $\mathbf{H}_d$ in Fig. 3.1, from a TSA at the transmit lens antenna, say tsa_t, to a TSA at the receive lens antenna, say tsa_r, is obtained as [39]

$$h_{\text{tsa}_t,\text{tsa}_r} = \varpi_d \sqrt{A_t A_r}, \quad \text{tsa}_t = 1, 2, \ldots, N_{\text{TSA}}, \quad \text{tsa}_r = 1, 2, \ldots, N_{\text{TSA}}, \quad (3.1)$$

where $\varpi_d = \alpha_d \beta_d$ is the product of small-scale fading ($\alpha_a \in \mathbb{C}$) and large-scale fading ($\beta_d \in \mathbb{R}$) of the dth path. The parameters α_d and β_d will be discussed later in the rest of the chapter. A_t and A_r denote respectively the transmit and receive lens aperture size. Here, A_t and A_r are directly related to N_{ray}, such that a bigger aperture means more rays arrive or depart a lens. The remaining $D - 1$ channel impulse responses
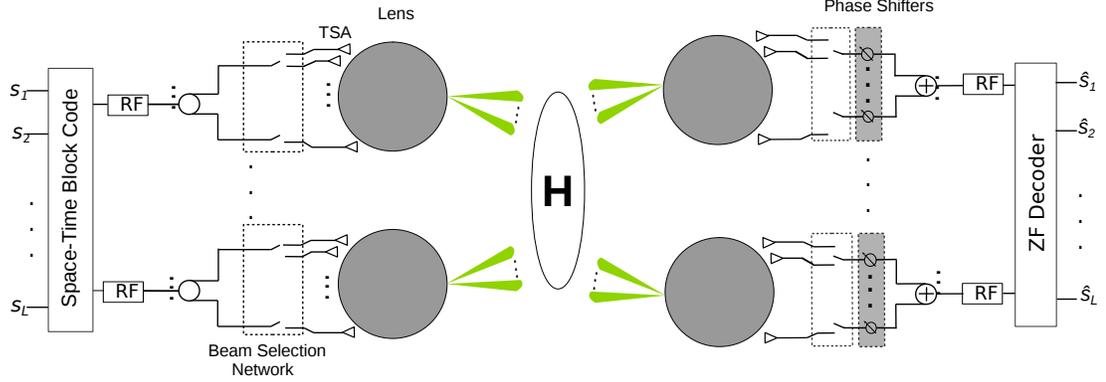
Figure 3.2   **The proposed RA-MIMO transceiver based on lens antennas.**

can be defined in the same fashion.

3.2.2   The Proposed RA-MIMO Architecture

As discussed earlier, hybrid beamforming systems compensate the high path loss through beamforming. However, they are not able to suppress the effect of small-scale fading and specially shadowing in NLoS channels. To obtain reliable transmission links, this chapter aims to propose a new mmWave MIMO structure. The proposed RA-MIMO is presented in Fig. 3.2. The proposed architecture has a fundamental property. In RA-MIMO, RF chains are separated and each of them is connected to a specific lens whereas those of the hybrid systems are spaced closely and connected to the same antenna array and lens, respectively. Therefore, our RA-MIMO architecture provides the degree of freedom that ensures the number of MIMO channel paths is dictated by the number of antennas and not by the mmWave channel.

3.2.3 System Parameters

We assume an outdoor[1] frequency flat fading[2] mmWave transceiver system composed of RA-MIMO at both the transmitter and the receiver as shown in Fig. 3.2. Further, it is assumed that the transceiver is located in an urban area where a large number of scattering objects are available. Therefore, sufficient independent paths are established. The transmitter and the receiver are equipped with N_t and N_r lens antennas, respectively. It is worth mentioning that each transmit reconfigure antenna steers N_r beams. Analogously, each receiver reconfigure antenna steers N_t beams.

Channel matrix

To determine the channel matrix of RA-MIMO, we notice that according to Fig. 3.2 there is only one channel between each transceiver lens antenna pair. That is, the n_tth antenna totally generates N_r beams, one beam for each receiver antenna. Analogously, the n_rth receive antenna totally generates N_t beams, one beam for each transmit antenna. It should be emphasized that D channels are established between each transceiver lens antenna pair. However, only one of the existing D channels is selected between each transceiver lens antenna pair. Therefore, (3.1) describes the channel at the TSA's outputs of a transceiver lens antenna pair. The channel between the n_tth lens antenna and the n_rth lens antenna at the TSA's outputs, by changing the

---

[1]For indoor communications in mmWave bands, analog systems are considered in which the transmitter and receiver are equipped with on RF chain [30–32]. In this case, transmission is done through one beam. Our proposed system can be used in indoor communications as well if only one RF chain and one beam are assumed.

[2]In the case of frequency selective fading, certain modulation schemes such as orthogonal frequency division multiplexing (OFDM) can be directly deployed in the proposed RA-MIMO to provide robustness to the fading.

notations of (3.1), becomes h_{n_t,n_r}. Eventually, the channel matrix $\mathbf{H}$ is expressed as

$$\mathbf{H} = [h_{n_t,n_r}]_{N_t \times N_r}. \tag{3.2}$$

It is important to note that in order to make sure $\mathbf{H}$ is full-rank, each channel should be used by only one transceiver pair. Hence, we have $N_t N_r \leq D$. In this chapter, it is assumed that the number of scattering objects (D) is sufficiently high. In the case that the number of scatterers is not high enough, the performance of RA-MIMO might be limited. One viable solution is that the antennas should be sufficiently separated such that multiple independent channels can be established by using one scatterer.

Reconfigured channel matrix

As outlined in [1], the proposed lens antenna can simultaneously change the phase of each beam. That is to say, in a single transmit/receive lens antenna, each TSA would be equipped by a phase-shifter and can change the phase of its signal independently. Although, we assume that only the receive lens antennas are equipped by phase-shifters. This is because in a mmWave MIMO system, the CSI associated to small-scale fading varies rapidly compared to the direction of each beam [78]. Since the steering direction changes slowly, this information can be useful at the transmitter. However, the fast varying CSI is of little value at the transmitter by the time it is estimated at the receiver and fed back. In our design, we assume that there is no CSI available at the transmitter. However, AoD and AoA information are respectively available at the transmitter. Mathematically, we represent role of the phase-shifters with matrix $\mathbf{G} = [g_{n_t,n_r}]_{N_t \times N_r}$. This matrix is called reconfigurable parameter matrix

of lens antenna. The parameter g_{n_t,n_r} is related to the n_tth beam of the n_rth antenna and is determined as

$$g_{n_t,n_r} = e^{j\theta_{n_t,n_r}}, \ n_t = 1, 2, \ldots, N_t, \ n_r = 1, 2, \ldots, N_r, \tag{3.3}$$

where θ_{n_t,n_r} denotes the phase of the corresponding phase-shifter at the receiver.

To realize the connection between the channel matrix and the reconfigurable parameter matrices, assume that a typical information signal, s, is sent from the n_rth beam of the n_tth antenna. The signal is then multiplied by the channel coefficient between the n_tth and the n_rth antenna, i.e., $s \times h_{n_t,n_r}$. At the receiver, it is multiplied by g_{n_t,n_r}. That is, the received signal is equal to $s \times h_{n_t,n_r} g_{n_t,n_r}$ without considering the noise term. This expression indicates that there is a one-to-one mapping among the entries of $\mathbf{G}$ and $\mathbf{H}$ at the receiver side. Hence, the Hadamard product can nicely describe this mapping. That is, the reconfigurability brings about a new matrix denoted by reconfigured channel matrix or $\mathbf{H}_g$, where

$$\mathbf{H}_g = \mathbf{H} \circ \mathbf{G}. \tag{3.4}$$

$\mathbf{H}_g = (\mathbf{h}_{g,1}, \mathbf{h}_{g,2}, \ldots, \mathbf{h}_{g,N_r})$ of size $N_t \times N_r$ with entries $\mathbf{h}_{g,n_r}$ of size $N_t \times 1$. Here, $\mathbf{h}_{g,n_r}$ denotes the reconfigured channel between the transmitter and the n_rth receiver antenna.

Signal and noise model

Now, we express the relationship between the transmit and receive antennas in baseband as

$$\mathbf{Y} = \sqrt{\frac{P_t}{N_t N_r}} \mathbf{X}(\mathbf{s})\mathbf{H}_g + \mathbf{Z}, \tag{3.5}$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_r})$ is the $T \times N_r$ received signal matrix. P_t denotes the transmit power. The factor $\frac{P_t}{N_t N_r}$ is per beam transmit power. More details about this factor will be provided in Section 3.3.3. T is the number of time slots in each block. $\mathbf{X}(\mathbf{s})$ is a $T \times N_t$ STBC matrix and $\mathbf{s} = (s_1, s_2, \ldots, s_L)^{\dagger}$ is the $L \times 1$ information signal vector, where its elements are drawn from constellation $\mathcal{A}$. $\mathbf{H}_g$ of size $N_t \times N_r$ is defined in (3.4). The elements of the $T \times N_r$ noise matrix, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{N_r})$, are modeled by independent identically distributed (i.i.d) complex Gaussian random variables with zero mean and σ^2 variance, i.e., $z_{t,n_r} \sim \mathcal{CN} = (0, \sigma^2)$, where $\sigma = \sqrt{\mathbb{E}\{|z_{t,n_r}|^2\}}$ for $t = 1, 2, \ldots, T$ and $n_r = 1, 2, \ldots, N_r$. The value of σ depends on the number of selected beams at each receive antennas. Compared to the traditional MIMO systems, (3.5) contains the reconfigured channel matrix ($\mathbf{H}_g$) rather than the channel matrix ($\mathbf{H}$). The parameters $\mathbf{G}$ and $\mathbf{X}(\mathbf{s})$ in (3.5) are flexible and should be designed.

It is worth mentioning that channel estimation is one of the important issues in implementing RA-MIMO. We divide channel estimation into two steps. At the first step, using beam search algorithms, the direction information (AoD and AoA) of each path is determined. Then, the small-scale fading and the large-scale fading are estimated. Since we assume that the lens antennas are sufficiently close, all D scatters between a typical transceiver can be seen by all antennas. This means if the channel between one transmit and one receiver antenna is estimated, the estimation is valid for

all transceiver pairs. Due to the directional transmission, finding accurate direction information is not trivial. Two main direction estimation algorithms in the mmWave band are i) exhaustive beam search [23] and ii) hierarchical beam search [79]. Both algorithms are extensively studied in hybrid beamforming systems with phase-shifters and are directly or indirectly applicable in RA-MIMO system.

Beam selection is another important issue for RA-MIMO. A proper beam selection method can increase diversity gain. Although, in this chapter, our focus is on diversity gain obtained from STBCs and phase-shifters. To this end, only $D = N_r N_t$ channels are considered to be available and each transceiver pair selects a channel randomly. Notice that all channels are estimated before. Further, it is assumed that the channels experience the same path loss. This assumption is almost firm since the channels are considered to be NLoS [67]. However, in the case that the number of channels is higher than $N_t N_r$ and each channel experiences different path loss, beam selection combined with STBCs can increase the diversity gain. That is, as $D > N_t N_r$, some or all transceiver antenna pairs have more than one option to select beam. So, an efficient beam selection algorithm will definitely lead to higher diversity gain. Some of the initial results on beam selection can be found in [80]. Channel estimation and beam selection are subject of future research.

In what follows, we analyse the performance of the proposed RA-MIMO considering the effect of small-scale fading and large-scale fading. The probability density function (PDF) of the small-scale fading and large-scale fading when they are superimposed is highly complicated [81, Eq. (6)]. Therefore, the analysis of the joint impact of small-scale fading and shadow fading will be intractable. Although a simplified expression for the PDF is studied, the obtained PDF is still complicated [82, Eq. (3)]

and the analysis will not be easy to follow. Hence, due to the complicated PDF, the small-scale fading and large-scale fading are evaluated separately.

3.3   Overcoming Small-Scale Fading

Each element of channel matrix $\mathbf{H}$ contains the effect of small-scale, large-scale fading and the transceiver lens gain. In this section, only small-scale fading is assumed and the effect of large-scale fading and the transceiver lens gain are dealt with in the next section. Hence, we assume $h_{n_r,n_t} = \alpha_d$, and small-scale fading is modeled by Rayleigh fading, i.e., $\alpha_d \sim \mathcal{CN}(0, 1)$[3]. The importance of suppressing the impact of small-scale fading is that the rapid and high fluctuations of the channel can cause unreliable handover decisions and transmission link failures [83].

3.3.1   Design of System Parameters

Design of θ_{n_t,n_r}

When the signal is received from multiple paths, the best technique to enhance the received SNR is to use maximum ratio combining (MRC) [84]. Utilizing the MRC technique requires weighing the signals proportional to the inverse of the channel coefficient amplitudes before combining. However, due to using phase-shifters in the RA-MIMO, this technique is not practical.

The alternative technique to improve the received SNR is the equal gain combining (EGC) [84]. This technique equally weights and co-phases the signals. Therefore, we

---

[3]In this chapter, to provide insight to the impact of the proposed system on the performance of mmWave NLoS channels, Rayleigh fading is assumed. The case of Nakagami fading which generalizes Rayleigh fading and approximates Rician fading is subject of the future work.

have

$$\theta_{n_t,n_r} = -\angle h_{n_t,n_r}, \quad n_t = 1, 2, \ldots, N_t, \quad n_r = 1, 2, \ldots, N_r. \tag{3.6}$$

Thus, $g_{n_t,n_r} = e^{-j\angle h_{n_t,n_r}}$, and $\mathbf{H}_g$ in (3.4) is a non-negative real-valued matrix while the channel matrix, $\mathbf{H}$, is complex-valued. In this section, we consider ideal phase-shifters, while digital phase-shifters are covered in Section 3.3.2. It is worth mentioning that in the case of wide-band transmission, a single value of a phase shift will not be valid across the whole band. To overcome this issue, one feasible solution is to deploy filter bank behind TSAs [85]. Next, phase of a received signal at the output of each filter is changed by using a phase-shifter. Then, the output signals are combined.

Design of the STBC Matrix X(s)

To attain full-diversity gain when CSI is not available at the transmitter, space-time block coding is considered as a promising method. STBCs have been designed for both real and complex channels [71]. Real STBCs are used for transmission over real-valued channels, and the complex STBCs are utilized for transmission over complex-valued channels. Note that the channel matrix, $\mathbf{H}$, in our system is complex-valued and naturally a complex STBC can be used. However, we propose compensating for the channel phase, as done in (3.6), and using a real STBC. Since real-valued rate-one orthogonal STBCs (OSTBCs) that provide the maximum diversity and simple symbol-by-symbol decoding exist for any number of transmit antennas, this will allow us to utilize a rate-one STBC and reduce the overall complexity by using linear receivers such as ZF. **Note that rate-one complex OSTBCs do no exist for more than two antennas [86].**

A set of real-valued square STBCs for $N_t = 2, 4$, and 8 have been designed in [86]. These STBCs are suitable candidates for the proposed RA-MIMO even though here we use them with complex constellation points. Two examples of rate-one $\mathbf{X}(\mathbf{s})$ for $N_t = 2$ and 4 are given by [86] as

$$\mathbf{X}(\mathbf{s}) = \begin{bmatrix} s_1 & s_2 \\ -s_2 & s_1 \end{bmatrix}, \tag{3.7}$$

and

$$\mathbf{X}(\mathbf{s}) = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ -s_2 & s_1 & -s_4 & s_3 \\ -s_3 & s_4 & s_1 & -s_2 \\ -s_4 & -s_3 & s_2 & s_1 \end{bmatrix}, \tag{3.8}$$

respectively. Note that while the codes are similar to the STBC in ([86], Eqs. (3) and (4)), their application in the proposed design is different. Other codes for any number of antennas can be found in [86]. All these codes can be represented as

$$\mathbf{X}(\mathbf{s}) = \mathbf{A}_1 s_1 + \mathbf{A}_2 s_2 + \cdots + \mathbf{A}_L s_L, \tag{3.9}$$

where $\mathbf{A}_\ell$ of size $T \times N_t$ is Hurwitz-Radon family of matrices which its entries are restricted to the set $\{-1,0,1\}$. In our design, the information symbols s_ℓ for $\ell = 1, 2, \ldots, L$ in (3.9) are drawn from a complex-valued constellation. Further, in [86],

it is proved that these matrices have the following properties:

$$\mathbf{A}_k^\dagger \mathbf{A}_k = \mathbf{I}_{N_t}, \qquad k = 1, 2, \ldots, L,$$

$$\mathbf{A}_k^\dagger \mathbf{A}_\ell = -\mathbf{A}_\ell \mathbf{A}_k^\dagger, \qquad 1 \leq k < \ell \leq L. \tag{3.10}$$

The most important observation about (3.10) is that for $\ell = k$ the matrices are orthogonal and for $\ell \neq k$ they are skew-symmetric. In what follows, we will use these properties to evaluate the STBCs in our RA-MIMO system.

Remark 1. As mentioned, there is a specific difference between the STBCs in [86] and the ones in this chapter. The STBC defined by (3.9) encodes **complex-valued symbols**, whereas [86] encodes **real-valued symbols**. Nevertheless, in both cases, the code rate is equal to one, and the codes facilitate low complexity decoding.

Remark 2. To encode complex-valued information symbols for $N_t = 2$ and 4, the OSTBC [87] and the Quasi-OSTBC (QOSTBC) [88] are well-suited rate-one codes. However, those codes are designed for a complex-valued channel matrix in which s and the conjugate of s are encoded to achieve full-diversity gain. Although the channel matrix $\mathbf{H}$ in our system is complex-valued, thanks to the lens antennas and phase compensation in (3.6), the complex-valued elements are converted into real-valued. That is, the reconfigure channel matrix in (3.4) is real-valued. Therefore, the conjugate of the information symbols is not required.

Remark 3. While real OSTBCs exist for any number of antennas, for more than four antennas, low complexity rate-one full-diversity complex STBCs are not available [86]. As such, in addition to low complexity and full-diversity, the rate advantage of the

proposed system will result in a high coding gain for a large number of transmit antennas.

3.3.2 Diversity and Error Probability Evaluation

This section studies the achievable diversity gain and derives an upper bound for error probability in the presence of digital phase-shifters. The limited resolution of digital phase-shifters can affect the performance of the proposed RA-MIMO system regarding diversity and coding gain. Full-diversity is the most important factor in designing STBCs because it dramatically reduces the error probability. On the other hand, coding gain is important since at high SNR regions a low coding gain can cause a big difference in error probability. Therefore, we first study the full-diversity and SER performance of the utilized STBCs with ideal phase-shifters. Then, we perform full-diversity and SER analysis for the digital phase-shifters.

Ideal phase-shifter

By ideal phase-shifter, we mean the phase-shifters have unlimited resolution and (3.6) holds. The following lemma demonstrates that under ideal phase-shifter the proposed RA-MIMO system achieves full-diversity.

Lemma 1. The proposed RA-MIMO system achieves full-diversity with linear receivers and ideal phase-shifters.

Proof. See Appendix A.1. □

Accordingly, the upper bound expression for the SER is given in the following theorem.

Theorem 1. In RA-MIMO system, the upper bound of the SER when using a ZF receiver with the cardinality of M is obtained as

$$\mathrm{P}(\hat{s}_\ell \to s_\ell) \leq \frac{M-1}{M} a^{-N_t N_r} \rho^{-N_t N_r}, \tag{3.11}$$

where $a = \frac{3}{2(M^2-1)}, \frac{\sin^2(\pi/M)}{2}$, and $\frac{3}{4(M-1)}$ for PAM, PSK, and square QAM constellations, respectively. Also, ρ denotes the average received SNR of each symbol which is given by

$$\rho_\ell = \frac{P_t}{N_t N_r} \frac{1}{(\boldsymbol{S}^{-1})_{\ell,\ell} \sigma^2} = \frac{P_t}{N_t N_r} \frac{\|\mathbf{h}_g\|^2}{\sigma^2}, \quad \ell = 1, 2, \cdots, L, \tag{3.12}$$

where $\boldsymbol{S}$ is defined in (A.5) and the second equality follows from (A.5) in Appendix A.1.

Proof. See Theorem 1 in [89]. □

Theorem 1 reveals that the upper bound is not affected by the reconfigured channel matrix. This means the proposed RA-MIMO reaches the maximum coding gain with the ZF receiver, which is usually sub-optimal for other systems. Consequently, minimum mean square error (MMSE) attains the maximum coding gain as well as a maximum likelihood (ML) receiver.

Digital phase-shifter

In real-time applications, limited resolution of digital phase-shifters becomes a serious challenge for achieving full-diversity gain. The reason is that the discrete resolution of the phase-shifter does not allow it to completely compensate the phase of channel

coefficients. When B-bit resolution phase-shifters are employed, the options to select a proper phase are constrained to be $\{0, \ldots, (2^B - 1)2\pi/2^B\}$. Therefore, the reconfigured channel matrix $\mathbf{H}_g$ given by (3.4) is no longer real. This would diminish the diversity gain and coding gain.

Lemma 2. For any number of encoded symbols L, the proposed RA-MIMO system can achieve full-diversity with a linear receiver and a B-bit resolution digital phase-shifter for $B > 1$. While providing full-diversity, small values of B may result in an ill-conditioned system.

Proof. See Appendix A.2. □

Theorem 2. The upper bound of the SER of RA-MIMO system using B-bit digital phase-shifters for $B = 2$ is given by

$$\mathrm{P}(\hat{s}_\ell \rightarrow s_\ell) < \frac{M-1}{M} a^{-N_t N_r} \rho^{-N_t N_r} C'^{-N_t N_r}, \tag{3.13}$$

where $C' = \frac{1}{C} \left(\frac{L}{L-1}\right)^{L-1}$ in which C is given by $\det(\bar{\bar{\boldsymbol{S}}}^{-1}) \geq C$ and $\bar{\bar{\boldsymbol{S}}}$ denotes the normalization of (A.8) presented in Appendix A.2.

Proof. See Appendix A.3. □

3.3.3 System Design Tradeoffs

In sub-6 GHz MIMO systems which are equipped with omni-directional antennas, the power is divided per antennas. As said in Section 3.2, the power in RA-MIMO is divided per beam. It is because in RA-MIMO each beam is steered by a single TSA. Since each TSA is an independent antenna, at the transmitter, power is divided by the number of all beams, i.e., $N_t N_r$. On the other hand, at the receiver, each independent

TSA adds an $i.i.d$ noise term to its own received signal. For instance, in a receive lens antenna with b beams, the received signal contains b $i.i.d$ noise terms. This is not true in sub-6 GHz MIMO systems. That is, in sub-6 GHz MIMO systems, a signal received through an omni-directional antenna, directional antennas, or reconfigurable antennas which add only one additive, white Gaussian noise (AWGN), to the received signal. Since the number of beams of each lens antenna equals the number of transmit lens antennas (N_t), increasing N_t will likely impose more noise terms which reduces the received SNR. This issue is shown by the following corollary.

Corollary 1. In RA-MIMO system, the average received SNR is given by

$$\rho = \frac{P_t}{N_t \sigma_n^2},\tag{3.14}$$

where σ_n^2 is variance of the noise at each TSA.

Proof. See Appendix A.4. $\square$

Remark 4. Although it seems that the value of N_r can affect the received SNR (See Eq. (3.5)), Corollary 1 indicates that only N_t has impact on the received SNR. This important finding leads to a tradeoff in designing RA-MIMO systems. That is, for various N_t, the proposed RA-MIMO will show different received SNR, which can lead to different coding gains. For instance, suppose two RA-MIMO systems achieve the same diversity gain with a different number of transmit and receive antennas, i.e., $N_t N_r = N_t' N_r'$. It is obvious that, for $N_t > N_t'$, we will have $N_r < N_r'$. Using (3.14), the second system attains a higher coding gain since it has less number of antennas compared to the first system.

3.4   Overcoming Shadowing

Large-scale fading contains two parts, path loss denoted by $\Upsilon^{-\frac{n}{2}}$ and shadowing denoted by β_d, e.g., $\beta'_d = \Upsilon^{-\frac{n}{2}}\beta_d$, where Υ is the distance between the transmitter and receiver, n is the path loss factor. More precisely, path loss is defined as the constant reduction at the received SNR as the distance between the transmitter and the receiver increases. The path loss factor highly depends on the propagation environment and the operational frequency. The values of the factor for different environments and operational frequencies at mmWave bands are obtained in [67]. Two common ways of overcoming path loss are increasing the transmit power and the transceiver antenna gain. It is stated that the energy focusing property of the lens acts like beamforming [77] which increases the received SNR by $\sqrt{A_t A_r}$ gain. Hence, it is supposed that the path loss is made up by the transmit power and the lenses. Therefore, we normalize the channel coefficient (3.1) by $\Upsilon^{-\frac{n}{2}}\sqrt{A_t A_r}$ which results in $h_{n_r,n_t} = \alpha_d\beta_d$. In what follows, we only concentrate on the impact of shadowing.

To focus on the impact of shadowing, we assume the impact of amplitude of small-scale fading is averaged out, i.e., $|\alpha_d| = 1$, and only consider the channel phase and shadowing, i.e., $h_{n_r,n_t} = e^{j\angle\alpha_d}\beta_d$. The obtained channel gain after applying (3.6), i.e., passing through the ideal phase-shifter, becomes $h_{g,n_r,n_t} = \beta_d$. Studies indicate that β_d^2 has a log-normal distribution with parameters μ and η^2 [67]. Applying logarithm operator on β_d^2 leads to a simple representation of shadowing such that $10\log_{10}\beta_d^2 \sim \mathcal{N}(\mu, \eta^2)$ in which $\mu = 0$ and η depends on the environment and operating frequency [67]. In the case of huge fluctuations, outage probability is supposed to be a proper criterion to evaluate performance of a system [67]. It should be noticed that the analysis is conducted for the ideal phase-shifters. The impact of using dig-

ital phase-shifters on the performance of RA-MIMO in the presence of shadowing is eliminated due to space limitations.

Theorem 3. The approximated outage probability for the proposed RA-MIMO under the influence of shadowing is given by

$$p_{\text{out}} \approx 1 - Q \left(\frac{10\log_{10}\left(\frac{\gamma\sigma^2 N_t N_r}{P_t}\right) - \mu'}{\eta'} \right), \tag{3.15}$$

where γ is the predefined threshold for outage probability. Also, μ' and η'^2 denote the mean and variance of the random variable $10\log_{10}\left(\|\mathbf{h}_g\|^2\right)$ which approximately has a normal distribution.

Proof. See Appendix A.5. $\square$

Remark 5. One cannot use traditional MIMO systems with omni-directional antennas and STBCs in mmWave bands due to high path loss. Hence, RA-MIMO provides an ability to benefit from traditional MIMO structures by using more directional gain antennas to both overcome path loss and shadowing and small-scale fading based on the proposed methodology. Further, one alternative may be using only one RF chain combined with equal gain combining. This causes feedback overhead and using high-speed switches and phase-shifters since the best beam should frequently be selected at the transmitter and receiver. Feedback overhead reduces the allocated time for data transmission and high-speed switches and phase-shifters operating as mmWave frequencies are extremely expensive. Moreover, in this chapter we aim to overcome both the small-scale fading and the large-scale fading. To this end, we have proposed a MIMO architecture in which several replicas of a transmit signal are received. This

enables the proposed MIMO system to simultaneously overcome the small-scale fading and the large-scale fading. However, in a system with one RF chain and equal gain combining it is impossible to simultaneously overcome the small-scale fading and large-scale fading.

Remark 6. The proposed RA-MIMO can be used in point-to-point wireless communications such as backhaul, fronthaul, and access links as well as in multi-user scenarios [16,17]. Moreover, it can be used for mmWave communications in the presence of LoS and NLoS channels. Especially, in LoS channels, the channel matrix becomes ill-conditioned and by adjusting the distance between the lenses which depends on the distance between the transmitter and receiver, a full-rank channel is obtained [90]. Further, transmit and receive diversity techniques as well as multiplexing techniques can be used in conjunction with the RA-MIMO system. RA-MIMO has a flexible structure which means the number of RF chains at the transmitter can be different from the receiver. Whereas, the existing hybrid beamforming systems require the same number of RF chains at transmitter and receiver ends [27–42]. That is to say, in hybrid beamforming systems, multiple data streams are sent simultaneously. Each data stream requires one RF chain at transmitter and one RF chain at receiver. Hence, the number of RF chains at transmitter and receiver should be the same. Last but not least, the lens antenna separation leads to the use of common scatterers to establish mutually orthogonal channels and therefore to achieve higher diversity gain or multiplexing gain [91].

3.5 Numerical Results

This section presents the numerical simulations for the proposed RA-MIMO in Fig. 3.2 along with the STBCs in (3.9) and ZF receivers. The communication channel is modeled as a Rayleigh fading channel where the channel parameters are modeled by a zero-mean and unit-variance Gaussian distribution. For shadowing, we consider an outdoor environment in an urban area, where the shadowing parameters are modeled by a log-normal distribution with a zero-mean and variance of 8 dB for each channel as modeled in [67]. Next, we note that there are three different types of hybrid beamforming systems mentioned in Section I. As such, these hybrid beamforming systems provide different beamforming gains that result in different throughputs. In particular, each hybrid beamforming transceiver has a beamforming gain, say G_t at transmitter or G_r at receiver. In the simulations, we set $G_t = G_r = 1$. That is, before the normalization, the received SNR is given by $\frac{P_t G_t G_r |h|^2}{N_t \sigma_n^2}$, where P_t and $|h|^2$ denote the transmit power and the channel gain, respectively, and σ_n^2 denotes the noise power. Further, due to a fixed distance between the transmitter and the receiver, the effect of pathloss is fixed and neglected. After normalization of the SNR by the beamforming gain $G_t G_r$, the same SNR is obtained for all three hybrid systems which is given by $\frac{P_t |h|^2}{N_t \sigma_n^2}$. On the other hand, our goal is to compare the proposed RA-MIMO with the hybrid beamforming systems in terms of the diversity gain. Since beamforming gain does not affect the diversity gain, it is rational to consider a normalized SNR. Hence, far from which hybrid beamforming system is considered for comparison, diversity gain is always the same. Throughout the simulations the term hybrid beamforming includes all the three types. Moreover, the STBCs in (3.7) and (3.8) are exploited in the hybrid beamforming systems for diversity gain comparison. Further, the described

channel conditions for the RA-MIMO are also used for the scenario of hybrid beam-
forming. To make a fair comparison, we normalize the SNR of RA-MIMO system
with its beamforming gain as well.

3.5.1 BER Performance Comparison

Fig. 3.3 shows the BER performance of the hybrid beamforming and the proposed
RA-MIMO systems versus SNR for a throughput of 4 bits/s/Hz. For a 2×2 MIMO
system, the symbols are drawn from 16-QAM and encoded by (3.7). The results
show that RA-MIMO achieves diversity of order 4, i.e., full-diversity gain, whereas,
the hybrid beamforming system shows a diversity of order 2 since each transmitted
signal is sent through one path at each time slot. Further, BER comparison is plotted
for RA-MIMO and hybrid beamforming systems for a 4×4 MIMO. The symbols are
encoded via (3.8). Likewise, the proposed RA-MIMO achieves diversity gain 16, i.e.,
full-diversity, while the hybrid beamforming gives diversity of order 4. Hence, the
BER is remarkably reduced. That is to say, transmission via the RA-MIMO is more
reliable than the hybrid beamforming. Another important result is that the proposed
RA-MIMO system can achieve same diversity gain as the hybrid beamforming system
using less RF chains. For instance, for a diversity of order 4, the hybrid beamforming
system requires 4 RF chains, while the RA-MIMO needs 2 RF chains. In general,
maximum diversity gain that can be achieved by the hybrid beamforming system is
N_{RF}, where N_{RF} denotes number of RF chains. Whereas, the proposed RA-MIMO
needs only $\lceil \sqrt{N_{\mathrm{RF}}} \rceil$ RF chains to achieve same or more diversity gain, where $\lceil . \rceil$
denotes the ceiling function.
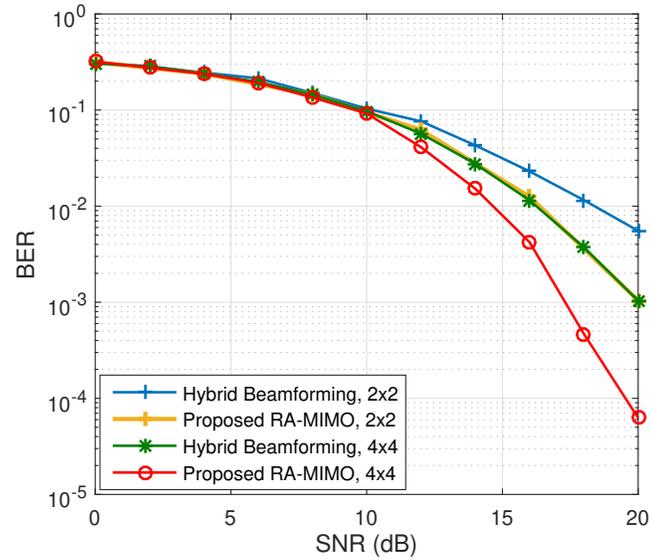
Figure 3.3 **BER performance of the hybrid beamforming system and the pro-posed RA-MIMO of size 2×2 and 4×4.**
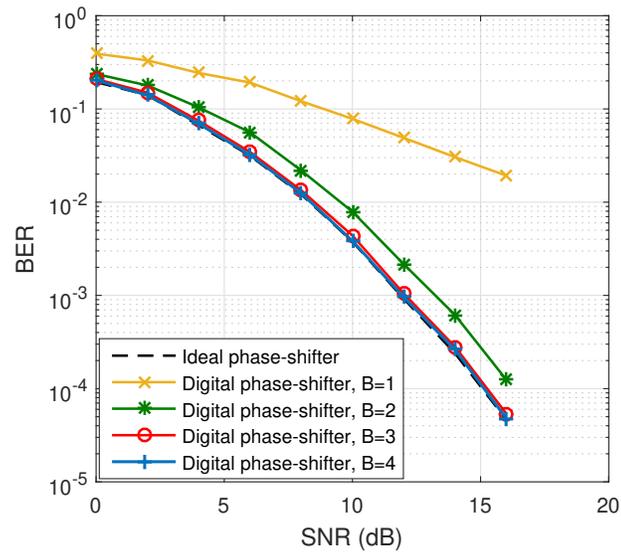
Figure 3.4: BER performance of 2×2 RA-MIMO versus SNR with ideal and digital phase-shifters.

3.5.2 BER Performance of Digital Phase-shifters

Fig. 3.4 shows BER curves for the RA-MIMO of dimension 2×2 with digital phase-shifters with $B = 1, 2, 3$, and 4. In this simulation, the STBC in (3.7) is used to encode two symbols ($L = 2$) over two time slots ($T = 2$). For comparison, the ideal phase-shifters are also considered. For $B = 1$, RA-MIMO does not provide full-diversity gain as shown in Appendix A.2. Simulation indicates that RA-MIMO achieves the full-diversity gain for the resolution of digital phase-shifters greater than 1 as well as ideal phase-shifters. This is consistent with Lemma 2 and Theorem 2. Further, as expected for low resolutions like ($B = 2$), RA-MIMO loses about 1 dB coding gain at BER=10^{-4} compared to that of the ideal phase-shifters. However, by increasing the number of bits in digital phase-shifters, the BER performance approaches to that of ideal phase-shifters such that for $B = 3$ and 4 the performances are approximately the same. The reason for a lower coding gain compared to the ideal phase-shifters is that in the case of digital phase-shifters due to the quatization error, the off-diagonal elements of the channel matrix in (A.8) are not zero. Thus, the channel's eigenvalues may not be equal as discussed in Theorem 2 which leads to a lower coding gain.

To show more comparison, we simulate RA-MIMO of dimension 4×4 by assuming the STBC in (3.8). Fig. 3.5 depicts the BER curves versus SNR of various resolutions for the digital phase-shifters. RA-MIMO with the ideal phase-shifters achieves full-diversity gain. However, as expected, the digital phase-shifters with $B = 2, 3$, and 4 achieve full-diversity gain, whereas, for $B = 1$, RA-MIMO does not provide full-diversity. It is worth noting that in the case of $B = 2$, the reconfigured channel becomes ill-conditioned as mentioned in Appendix A.2. This would severely degrade the coding gain of the system. Further, for $B = 3$ and 4, RA-MIMO with digital
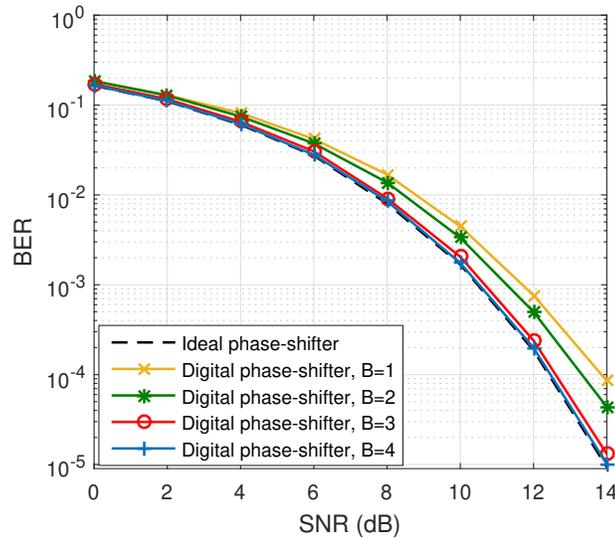
Figure 3.5: BER performance of 4×4 **RA-MIMO versus SNR with ideal and digital phase-shifters.**

phase-shifters approximately attains the same coding gain as that of ideal phase-shifters.

3.5.3 System Design Tradeoff

Fig. 3.6 reveals the impact of the antenna configuration on the BER performance of RA-MIMO. As it is mentioned in Section 3.3.3, N_t scales the received SNR such that increasing the N_t diminishes the SNR. To investigate this, we assume 4 different RA-MIMO antenna configurations 8×1, 4×2, 2×4, and 1×8. The last RA-MIMO uses a single-input multiple-output structure (SIMO) which is a special case of RA-MIMO. For Fig. 3.6 we conclude that all 4 RA-MIMO configurations achieve full-diversity gain but different coding gains. This verifies Corollary 1.
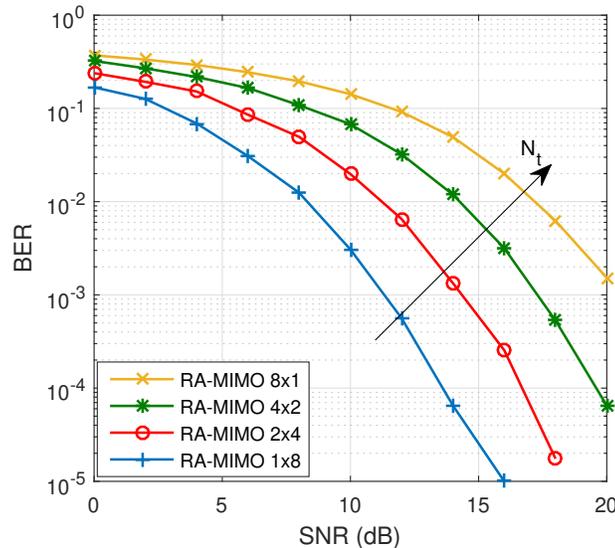
Figure 3.6 **BER performance for different antenna configurations.**

3.5.4 Outage Probability

In Fig. 3.7, the outage probability performance is illustrated for a hybrid beamforming system with two RF chains and different dimensions of RA-MIMO system. For the hybrid beamforming system, it is assumed that the transceiver is equipped with two RF chains, and two streams are simultaneously sent over two NLoS channels shown by "Hybrid Beamforming 2×2". For RA-MIMO, three configurations 2×2, 2×3, and 2×4 are considered. The threshold γ is set to 0 dB. As shown, the hybrid beamforming system provides the highest outage probability which is the worst performance. That is, using only one NLoS channel for transmission of a data stream will bring up frequent fluctuations which increases outage probability. Whereas, RA-MIMO is able to reduce outage probability when more number of NLoS channels are utilized. At SNR 40 dB, outage probability for the hybrid beamforming system is about 10^{-1}, while that of RA-MIMO for 2×2, 2×3, and 2×4 are 10^{-2}, 10^{-3}, and 10^{-5}, respectively.
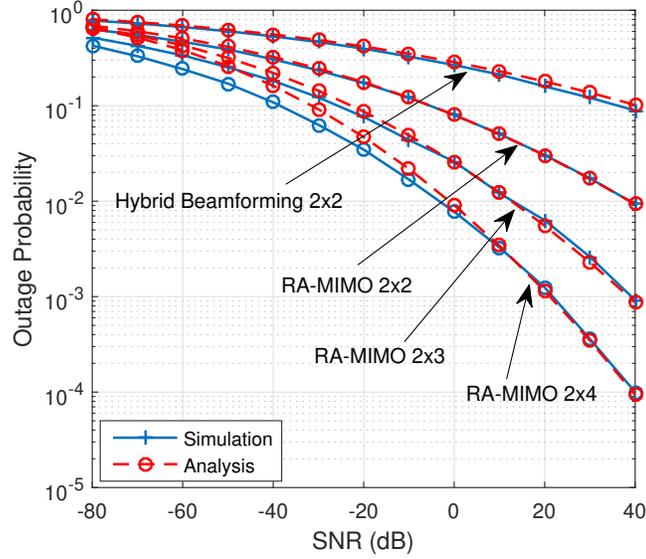
Figure 3.7: Comparison of outage probability between a hybrid beamforming system and various RA-MIMO configurations.

In addition, the derived expression for outage probability in Theorem 3 is also plotted. The simulation indicates that when the exact PDF of random variable v is replaced by an approximated normal distribution, the obtained outage probability is accurately fitted to the actual value of outage probability. By doing exhaustive search over various values, the log-normal parameters μ' and η' are obtained as follows. For RA-MIMO of dimensions 2×2, 2×3, and 2×4, the values of μ' are respectively given as 15, 16.9, and 17.9 and those of η' are respectively obtained as 9.7, 7.8, and 6.7 dB. Further, the parameters of the hybrid beamforming system is obtained as $\mu' = 8$ and $\eta' = 13$. It is revealed that by increasing the dimension of RA-MIMO μ' rises and η' diminishes.

3.6   Summary and Conclusion

In this chapter, an RA-MIMO system for mmWave wireless communications is proposed. RA-MIMO is inspired from the lens antennas and its aim is to combat small-scale fading and shadowing at mmWave frequencies. First of all, the channel matrix for the RA-MIMO is derived. Then, by using the lens antennas along with ideal phase-shifters, the complex-valued channel matrix is converted to the real-valued one. Further, RA-MIMO achieves full-diversity when complex-valued STBCs are utilized. Next, studying a practical scenario, it is shown that even with the digital phase-shifters, the system achieves full-diversity when the phase-shifters' resolution is greater than one-bit at the cost of losing coding gain. Upper bounds are derived for the SER performance in both ideal and digital phase-shifters. By calculating the received SNR with respect to N_t, a tradeoff between the number of transmit antennas and coding gain is found. Lastly, the outage probability of RA-MIMO in the presence of shadowing is studied. The numerical results confirm our analytical findings.

CHAPTER 4:

NEW MULTIPLE ACCESS TECHNIQUES WITH LENS-BASED MMWAVE ANTENNA

This chapter aims to realize two new multiple access techniques based on millimeter-wave reconfigurable antenna architectures introduced in Chapter 3. To this end, first, a new multiple access technique is proposed which is called reconfigurable antenna multiple access (RAMA) that transmits only each user's intended signal at the same time/frequency/code. This property makes RAMA an inter-user interference-free technique. Second, to support a large number of groups of users then we integrate RAMA into NOMA. The proposed technique, named reconfigurable antenna NOMA (RA-NOMA), divides the users with respect to their AoDs and channel gains. Users with different AoDs and comparable channel gains are served via RAMA while users with the same AoDs but different channel gains are served via NOMA. This technique results in the independence of the number of radio frequency chains from the number of NOMA groups. Further, we derive the feasibility conditions and show that the power allocation for RA-NOMA is a convex problem. We then derive the maximum achievable sum-rate of RA-NOMA. The numerical results demonstrate that RAMA outperforms NOMA in terms of individual user rate and sum-rate. Also, RA-NOMA shows higher sum-rate compared to OMA, NOMA, and NOMA-OMA.

4.1   Introduction

Recall that, in addition to the mmWave beamforming systems, non-orthogonal multiple access (NOMA) has been also considered as another promising enabling technique for 5G to enhance spectral efficiency in multi-user scenarios [56,57]. In order to serve more users in 5G wireless communications, recently, the integration of NOMA in mmWave systems, i.e., mmWave-NOMA, has been studied [92–94]. MmWave-NOMA systems are completely discussed in the next chapter. Further, in the previous chapter, it is mentioned that the reconfigurable antennas can support the transmission of multiple radiation beams with one RF chain [1, 13, 74]. Given that the limitation on the RF circuitry of the antenna [1] results in the division of the transmitted power amongst the beams, the current state-of-the-art in mmWave-NOMA would not operate efficiently in such a setting.

4.1.1   Contributions

To enhance the performance of multiple access schemes in the mmWave band and also overcome this fundamental limit for reconfigurable antennas, RAMA and RA-NOMA techniques are proposed. The major contribution of this chapter are listed as follows.

- We propose a fresh multiple access technique for mmWave reconfigurable antenna systems which is called RAMA [16]. Since RAMA aims to transmit only the intended signal of each user, SC and SIC are not required.

- We compare sum-rate of RAMA and NOMA. It is shown that for two users, RAMA always achieves higher sum-rate. We compare the RAMA and NOMA in terms of their achievable maximum sum-rate. It is shown that RAMA can operate in two cases. In the first case, it is assumed that only angle of depar-

tures (AoDs) are available at the BS. Hence, users are served via fixed power allocation. Then, it is assumed that the BS knows full CSI, i.e., channel gains and AoDs. In this case, dynamic power allocation strategy is suggested which leads to higher sum-rate.

- Although RAMA achieves a higher sum-rate than NOMA, its operation is limited to only one user per beam. To overcome this limit, we propose a new multiuser technique which integrates RAMA and NOMA, and is referred to as RA-NOMA.

- In RA-NOMA, each antenna element is connected to all RF chains. Users are grouped into two different sets based on the AoDs and the users' channel gain. The users with different AoDs and comparable channel gains are supported via RAMA while the users with the same AoDs but different channel gains are served via NOMA. This technique introduces a new degree of freedom for grouping the users without adding RF chains.

- We show that RA-NOMA can be considered as a multi-beam NOMA scenario. Then, the power allocation problem is transformed into a convex problem over one beam. Further, the feasibility conditions of the power allocation problem are assessed.

Numerical results reveal that in the case of one RF chain the proposed RAMA outperforms OMA and NOMA in terms of individual user rate and sum-rate. Also, for higher number of RF chains, the proposed RA-NOMA achieves higher sum-rate than OMA, NOMA and the combination of RAMA with OMA.

The chapter is organized as follows. In Section 4.2, the system model is described.

Section 4.3.1 proposes RAMA. Section 4.4 compares the sum-rate between RAMA and NOMA. Section 4.5 proposes RA-NOMA technique. Section 4.6 presents the power allocation in the downlink of RA-NOMA. In Section 4.7, simulation results are presented. Finally, Section 4.8 concludes the chapter.

Notations: Hereafter, $j = \sqrt{-1}$. Also, $\mathbb{E}[\cdot]$ and $|\cdot|$ denote the expected value and amplitude value of $(\cdot)$, respectively. Small letters and bold letters designate scalars and vectors, respectively. Further, $\|\mathbf{x}\|_1 = \sum_i |x_i|$ denotes the 1-norm of vector $\mathbf{x}$ and superscript $(\cdot)^\dagger$ denotes the transpose operator.

4.2   System Model

We consider the downlink of a single mmWave narrow band communication cell with a base station (BS) in the center that serves K' users. The users are uniformly distributed. Further, the BS is equipped with the lens-based mmWave reconfigurable antenna and each user uses a single omni-directional antenna.

A lens-based mmWave reconfigurable antenna comprises of the following blocks: one RF chain, one splitter, one switching network, N_{TSA} tapered slot antennas (TSAs), and one spherical lens located in front of the TSAs. The system is shown in Fig. 4.1. The reconfigurable antenna is supposed to simultaneously generate N_{B} beams when only one RF chain is deployed. This is done via the switching network by selecting N_{B} beams of N_{TSA} options. Further, the reconfigurable antenna steers independent beams. That is, the antenna can independently manipulate the phase and magnitude of each beam. Therefore, steering multiple independent beams is the main property of the reconfigurable antennas. This important property is used to design a new multiple access technique for multiuser networks.

The single LoS channel for User k is given by $\mathbf{h}_k = h_k \mathbf{a}(\theta_k)$, where $h_k = \alpha_k d_k^{-\frac{v}{2}} \in \mathbb{C}$
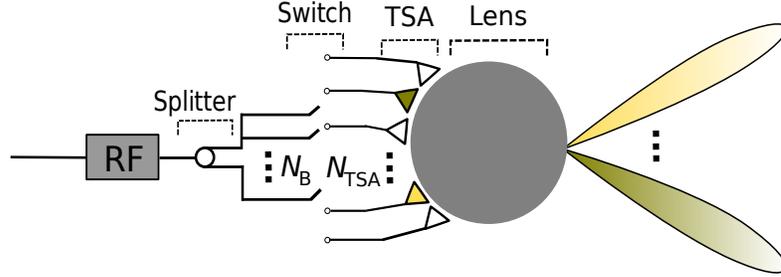
Figure 4.1: Schematic of the reconfigurable antenna steering multiple beams. The antenna is composed of a spherical lens fed with a number of tapered slot antenna feeds. Each feed generates a beam in a given direction in the far field [1].

is a composite model which consists of the small scale fading α_k with zero-mean and unit-variance and the large scale path loss $d_k^{-\frac{v}{2}}$ in which d_k is the path distance between the BS and the user and v is the path loss factor. $\mathbf{a}(\theta_k)$ denotes the antenna array steering vector of the BS, and θ_k denotes the azimuth AoD.

From Chapter 3, we recall that the spherical lens operates as a passive phase shifter network that shifts the phase of an array with proper delays [77]. Hence, the effective channel between the BS and each user is given by

$$\mathbf{a}^*(\theta_k)\mathbf{h}_k = h_k, \tag{4.1}$$

Therefore, the channel gain is given as $|h_k|^2 = |\alpha_k|^2 d_k^{-v}$. We sort the users regarding their distance from the BS rather than their channel gain, i.e.,

$$d_1 \geq d_2 \geq \cdots \geq d_{K'}. \tag{4.2}$$

It is shown by [95,96] that the small scale fading α_k weakly changes the distribution of the variable $|h_k|^2$. So, this motivates us to sort the users based upon the distances. Further, the k user is defined by User k.

4.3 The Proposed Reconfigurable Antenna Multiple Access

As it is mentioned, there are K' users to be served. Without of loss generality, it is assumed that among the K' users, K users are located in distinct directions. In Section 4.5 we propose another technique which serves users far from their direction. So, in this section we focus only on those users that are not located at same direction. We show that when the K users are simultaneously served via our proposed RAMA, the sum-rate is higher than NOMA. To better explain the advantages of the proposed RAMA, first we review NOMA. Then, NOMA is implemented by the reconfigurable antenna. While the proposed technique and the results can easily be extended to an arbitrary number of users, for the sake of simplicity, only two users are considered. Hence, $K = 2$.

4.3.1 NOMA with Lens-based Reconfigurable Antennas

Suppose that the BS is equipped with the reconfigurable antenna described in Section 4.2. Our aim is to simultaneously serve Users 1 and 2 by using the reconfigurable antenna and NOMA. The reconfigurable antenna steers two beams by feeding two TSAs. The superposition coding of s_k with allocated power p_k is defined in (2.4). Users 1 and 2 receive the following signals as

$$
\begin{cases}
y_1 = \frac{1}{\sqrt{2}} h_1 x + n_1 \equiv h_1 \left(\frac{\sqrt{p_1}}{\sqrt{2}} s_1 + \frac{\sqrt{p_2}}{\sqrt{2}} s_2 \right) + n_1, \\
y_2 = \frac{1}{\sqrt{2}} h_2 x + n_2 \equiv h_2 \left(\frac{\sqrt{p_1}}{\sqrt{2}} s_1 + \frac{\sqrt{p_2}}{\sqrt{2}} s_2 \right) + n_2,
\end{cases}
\tag{4.3}
$$

respectively, where factor $\frac{1}{\sqrt{2}}$ is due to the splitter in the reconfigurable antennas. Note that the splitter splits the power equally between the two beams which is shown in Fig. 4.1. Thus, each user receives only a portion of the power of the transmitted signal x which does not happen in NOMA when the BS uses omni-directional antenna. An error-free SIC process results in the achievable rate for each user as

$$
\begin{cases}
\bar{R}_1^N = \log_2(1 + \frac{\rho_1|h_1|^2}{2}), \\
\bar{R}_2^N = \log_2(1 + \frac{\frac{1}{2}\rho_2|h_2|^2}{\frac{1}{2}\rho_1|h_2|^2+1}) \equiv \log_2(1 + \frac{\rho_2|h_2|^2}{\rho_1|h_2|^2+2}).
\end{cases}
\tag{4.4}
$$

It is clear that the signal-to-interference-plus-noise ratio (SINR) for User 2 and the SNR for User 1 in (4.4) are less than those of NOMA given in (2.6). As a result, combining NOMA with reconfigurable antennas will reduce the achievable user rate when considering the same channel gains, power allocation, and noise power.

In brief, when users are not aligned in the same direction, the reconfigurable antennas cannot harvest the benefits of NOMA due to the power division. Whereas, when users are located on the same direction, the reconfigurable antenna steers only one beam to serve users which means power division is not required.

4.3.2 The Proposed Reconfigurable Antenna Multiple Access

The derived rate expressions in (4.4) reveal that the reconfigurable antenna amplifies the noise power without any impact on the interference. Since the power amplification is inevitable, our main objective is to suppress the inter-user interference. Toward this goal, we aim to transmit only the intended signal for each user at the same time/frequency/code blocks. The intended signals for Users 1 and 2 are s_1 and s_2, respectively. Let us assume that s_k are drawn from a phase shift keying (PSK)
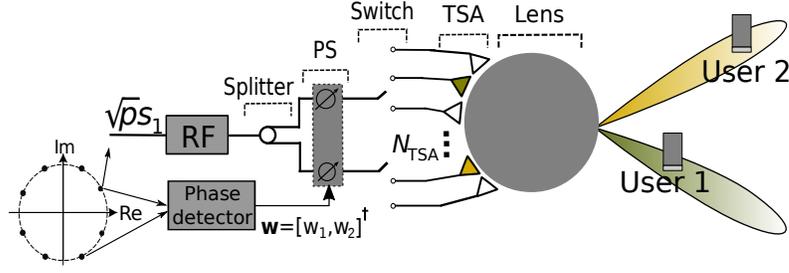
Figure 4.2   Two multiple access techniques for the reconfigurable antenna: (a) NOMA and (b) the proposed RAMA.

constellation and $\mathbb{E}[|s_k|^2] = 1$. Accordingly, s_1 and s_2 is expressed in terms of s_1 as

$$\mathbf{s} = \mathbf{w}s_1, \tag{4.5}$$

where $\mathbf{s} = [s_1, s_2]^\dagger$ and $\mathbf{w} = [1, e^{j\Delta\phi}]^\dagger$ where $\Delta\phi$ denotes the difference between the phases of s_1 and s_2. Similar to (2.4), the power of the transmitted signal, x, is assumed to be p. It is worth mentioning that the phase-shifters should change the phase symbol-bu-symbol. In practice, such phase-shifters may not exist or be very expensive. An alternative solution is to connect a fast switch to multiple phase-shifters. Hence, based on the difference between the phases, the proper phase-shifter is selected.

Unlike NOMA, in RAMA only one of the signals, say s_1, is upconverted by the RF chain block and the whole power p is allocated to that signal before the power division step. Therefore, x is given by

$$x = \sqrt{p}s_1. \tag{4.6}$$

The proposed multiple access technique for the signal in (4.6) is shown in Fig. 4.2.

The phase detector block calculates the phase difference between s_1 and s_2, $e^{j\Delta\phi}$. Moreover, as shown in Fig. 4.2, the beam selection network selects two TSA feeds regarding the AoDs. For simplicity, we call them TSA 1 and TSA 2. The network splits the power equally between TSA 1 and TSA 2. That is, the signal of TSAs 1 and 2 is given by $\frac{\sqrt{p}}{\sqrt{2}}s_1$.

The signal intended for TSA 1 is the desired signal for User 1. To transmit the signal intended for User 2 via TSA 2 we take advantage of the reconfigurable antennas. Thanks to the properties of the reconfigurable antennas, the signal at each TSA can independently be rotated with an arbitrary angle. Using this property, the beam selection network multiplies the signal in corresponding to TSA 2 with $e^{j\Delta\phi}$. This results in the transmitted signal via TSA 2 to be $\frac{\sqrt{p}}{\sqrt{2}}s_1 e^{j\Delta\phi} = \frac{\sqrt{p}}{\sqrt{2}}s_2$ which is the desired signal for User 2. Since transmission is directional in mmWave bands, each user receives only its intended signal as

$$\begin{cases} y_1 = \frac{\sqrt{p}}{\sqrt{2}}h_1 s_1 + n_1, \\ y_2 = \frac{\sqrt{p}}{\sqrt{2}}h_2 s_2 + n_2. \end{cases} \tag{4.7}$$

It is noteworthy that, here, we assume that there is no interference which is imposed from the signal intended for User 1 on User 2 and vice versa. This assumption is very well justified since the structure of the proposed lens based slotted reconfigurable antenna results in highly directional beams with very limited sidelobes [74]. Moreover, due to significant pathloss and shadowing at mmWave frequencies we do not expect the signals from the sidelobes to reach the unintended users. We also highlight that in contrast to NOMA the SIC process are not required at the receiver.

The achievable rate of RAMA for each user under equal power allocation is ob-

tained as

$$
\begin{cases}
R_1 = \log_2(1 + \rho_1 |h_1|^2), \\
R_2 = \log_2(1 + \rho_2 |h_2|^2),
\end{cases}
\tag{4.8}
$$

where $\rho_k = \frac{p}{2\sigma^2}$. Compared to (2.6), RAMA is able to remove the inter-user interference at the cost of adding noise. Also, compared to (4.4), in (4.8) the noise power remains unchanged, but the interference is eliminated.

4.4 Rate Analysis

In (4.8), the users are allocated equal powers and only AoDs are available at the BS, i.e., partial CSI. In this case, the users' power is fixed which means the users are served with fixed power-RAMA (F-RAMA). Another possible case is to use dynamic power allocation. That is, the users are served with unequal power levels regarding their channel gain. In this case, full CSI is available at the BS. We call this technique dynamic power-RAMA (D-RAMA). To implement D-RAMA technique, the splitter in Fig. 4.2 divides the power dynamically among the TSAs [97]. By definition, sum-rate for NOMA and F-RAMA are $R_{\text{sum}}^N = R_1^N + R_2^N$ and $R_{\text{sum}} = R_1 + R_2$, respectively.

4.4.1 Rate Comparison for F-RAMA

Case I: $|h_1|^2 = |h_2|^2$. This case is called a symmetric channel [56]. That is, the two users have the same SNR. In this case, F-RAMA always achieves higher sum-rate than NOMA. To show this, we calculate the sum-rate for NOMA and F-RAMA. For

NOMA, the sum-rate can be calculated as

$$
\begin{aligned}
R_{\text{sum}}^{N} &= \log_2\big(1 + \frac{p_1|h_1|^2}{\sigma^2}\big) + \log_2\big(1 + \frac{p_2|h_2|^2/\sigma^2}{p_1|h_2|^2/\sigma^2 + 1}\big) \\
&\overset{(a)}{=} \log_2\bigg(\big(1 + \frac{p_1|h_1|^2}{\sigma^2}\big)\big(1 + \frac{p_2|h_2|^2/\sigma^2}{p_1|h_2|^2/\sigma^2 + 1}\big)\bigg) \\
&= \log_2\big(1 + \frac{p_1|h_1|^2}{\sigma^2} + \frac{p_2|h_2|^2}{\sigma^2}\big) \\
&\overset{(b)}{=} \log_2\big(1 + \frac{p|h|^2}{\sigma^2}\big).
\end{aligned}
\tag{4.9}
$$

The (a) follows from $\log_2 a + \log_2 b = \log_2 ab$ and the (b) follows the assumption that $|h|^2 = |h_1|^2 = |h_2|^2$ and $p_1 + p_2 = p$.

Also, for RAMA-I, we follow the same steps as in NOMA. Hence, it is obtained as

$$
R_{\text{sum}} = \log_2\big(1 + \frac{p|h|^2}{\sigma^2} + \frac{p^2|h|^4}{4\sigma^4}\big).
\tag{4.10}
$$

Since $p^2|h|^4/4\sigma^4 \geq 0$, it gives $R_{\text{sum}}^{N} \leq R_{\text{sum}}$.

Case II: $|h_1|^2 \geq |h_2|^2$. Here, this case is called an asymmetric channel [56]. That is, we assume that the channel gain for User 1 is stronger than User 2. It can be shown that for asymmetric channels, F-RAMA achieves higher sum-rate than NOMA when the power is properly allocated for Users 1 and 2[1]. To proof this claim, we have

$$
\begin{aligned}
R_{\text{sum}}^{N} &= \log_2(1 + \frac{p_1|h_1|^2}{\sigma^2}) + \log_2(1 + \frac{p_2|h_2|^2/\sigma^2}{p_1|h_2|^2/\sigma^2 + 1}) \\
&\overset{(a)}{=} \log_2\big((1 + \frac{p_1|h_1|^2}{\sigma^2})(1 + \frac{p_2|h_2|^2/\sigma^2}{p_1|h_2|^2/\sigma^2 + 1})\big) \\
&\overset{(b)}{\leq} \log_2\big((1 + \frac{p_1|h_1|^2}{\sigma^2})(1 + \frac{p_2|h_2|^2}{\sigma^2})\big),
\end{aligned}
\tag{4.11}
$$

---

[1]To achieve user fairness in NOMA, when $|h_1|^2 \geq |h_2|^2$ we have $p_1 \leq p_2$ [56].

where p_1 and p_2 are allocated power for Users 1 and 2, respectively. Also, the (a) follows from $\log_2 a + \log_2 b = \log_2 ab$, and the (b) is a result of $p_1|h_2|^2/\sigma^2 > 0$. Using (4.11) and (4.8), the inequality $R_{\text{sum}}^N \leq R_{\text{sum}}$ holds when the following condition follows

$$(1 + \frac{p_1|h_1|^2}{\sigma^2})(1 + \frac{p_2|h_2|^2}{\sigma^2}) \leq (1 + \frac{p|h_1|^2}{2\sigma^2})(1 + \frac{p|h_2|^2}{2\sigma^2}). \qquad (4.12)$$

Obviously, for $p_1/p \in (0, 0.5]$ the inequality holds which indicates that User 1 should have lower power than User 2. Although this range is not tight, it gives a considerable insight. This result implies that with proper power allocation in NOMA, F-RAMA attains higher sum-rate. However, F-RAMA may not achieve user fairness when channel gain of one of the users is significantly greater than that of other user. In this case, the allocated power for user with strong channel gain should be far less than other user and equal power allocation would not lead to user fairness.

It is straightforward to show that D-RAMA achieves higher sum-rate than NOMA irrespective of their channel condition. That is, $R_{\text{sum}}^N \leq R_{\text{sum}}$. Further, D-RAMA considers user fairness the same as NOMA.

4.4.2 The Following Remarks Are in Order:

Remark 7. While NOMA in Subsection 4.3.1 and D-RAMA require full CSI, i.e., the channel gains and angle information, F-RAMA only needs AoDs which reduces system overhead. In highly dense networks where the number of users is much larger than two, to avoid more overhead, it is reasonable to serve users with F-RAMA.

Remark 8. F-RAMA implemented by PSK constellation is very simple in practice. The beam selection network divides the power equally by using a simple power divider

and the received signal is interference-free. The later is also preserved for D-RAMA. Indeed, F-RAMA with quadrature amplitude modulation (QAM) constellations is operational if power is divided properly.

Remark 9. F-RAMA may not achieve user fairness when channel gain of one of the users is significantly greater than that of other user. In this case, the allocated power for the strong user should be far less than the weak user and equal power allocation would not lead to user fairness.

Remark 10. D-RAMA needs an optimal power allocation strategy. One feasible strategy is to adopt power allocation in NOMA. However, this strategy may not be efficient for D-RAMA. It is because NOMA considers the interference and the minimum achievable rate for each user to allocate the power [56]. Whereas, for D-RAMA the interference is removed which leads to designing a suitable power allocation strategy.

In practice, implementation of D-RAMA requires complicated hardware design which is not trivial. Specially, in mmWave frequencies, the active components in a dynamic power divider are expensive and energy consuming. To ease the implementation issue and have a cost-effective design, hereafter, we focus on the F-RAMA technique. Also, the term RAMA is used interchangeably with F-RAMA.

4.5   The Proposed RA-NOMA Structure

Recall that RAMA is only applicable to users with different AoDs and in directional transmission, NOMA is mostly applied for users with the same AoDs but different channel gains. This points out that the two techniques can be used to serve two different group of users. Hence, there is a unique opportunity to integrate RAMA

with NOMA, i.e., RA-NOMA, to serve far more users with a scalable structure.

Toward this goal, we divide the users into two groups based on their AoDs and channel gains as: 1) RAMA and 2) NOMA groups. The users in the RAMA group are assumed to have distinctive AoDs but close channel gains. The users within the NOMA group are assumed to have the same AoDs but different channel gains. Considering the same AoDs or channel gains is optimistic. However, users with close AoDs or channel gains with high probability are available in 5G networks in which users are densely deployed [98]. Fig. 4.3 depicts a potential configuration for RA-NOMA users. The users with the same color belong to the first group, e.g., red-colored users represent a group of RAMA users. Further, those users covered by the same beam represent the second group, or NOMA. For instance, see the users located within the area covered by yellow-colored beam in Fig. 4.3.

In our proposed design, we assume that the splitters have fixed number of outputs with equal power. Therefore, the number of beams supporting the NOMA groups is fixed. We assume that the number of beams, i.e., the NOMA groups, is denoted as N_B. The number of RAMA groups can be variable which equals to the number of RF chains. Therefore, N_{RF} is the number of users supported by each beam. Hence, total number of the served users via RA-NOMA is $N_B \times N_{\mathrm{RF}}$.

The transmission power of the ith RF chain, $i = 1, \ldots, N_{\mathrm{RF}}$, i.e., total allocated power to the ith group of RAMA users, is assumed to be p_i. αp_i denotes the allocated power to each user in the ith RAMA group. Hence, the allocated power to each user of the ith RAMA group becomes

$$\alpha p_i = \frac{p_i}{N_B}, \quad i = 1, \ldots N_{\mathrm{RF}}. \tag{4.13}$$
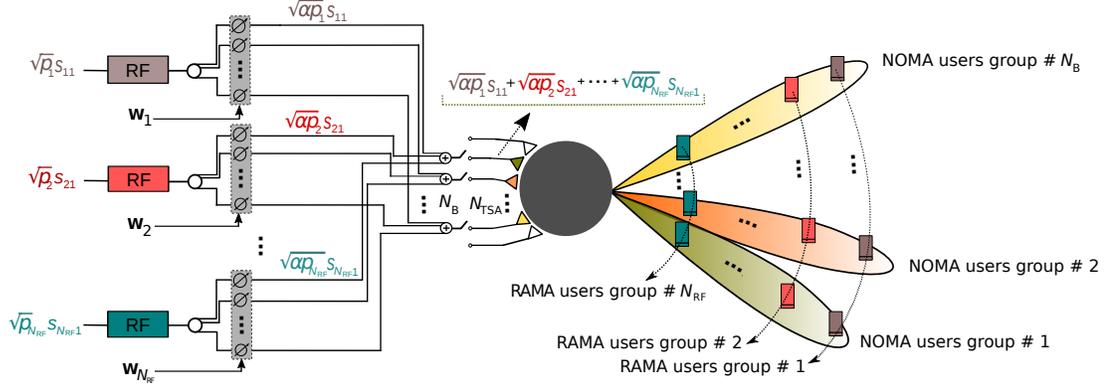
Figure 4.3 **The proposed RA-NOMA structure. The users are grouped into RAMA users and NOMA users. The number of RAMA users groups is equal to the number of RF chains and the number of NOMA users groups is the same as the number of beams.**

We let s_{i1} denote the intended signal for the ith group of RAMA users before the ith RF chain. At the ith phase shifter, the vector $\mathbf{w}_i$ is designed based on (4.5), such that we have

$$\mathbf{w}_i = [1, e^{j\Delta\theta_{i2}}, \ldots, e^{j\Delta\theta_{iN_B}}]^\dagger, \tag{4.14}$$

and

$$\mathbf{s}_i = \mathbf{w}_i s_{i1}, \tag{4.15}$$

where, $\mathbf{s}_i = [s_{i1}, \ldots, s_{iN_B}]^\dagger$ contains the intended signals for the users in the ith RAMA group. Considering the allocated power of each NOMA group, the superposition coded signal of the kth beam is given by

$$\sum_{i=1}^{N_{\text{RF}}} \sqrt{\alpha p_i} s_{ik} = \sqrt{\alpha p_1} s_{1k} + \cdots + \sqrt{\alpha p_{N_k}} s_{N_k k}. \tag{4.16}$$

Let h_{ik} refer to the channel gain of a user in the ith RAMA group in the kth

beam and we refer to this user as User (i,k). In the kth beam or equally the kth NOMA group, without loss of generality, we assume the channels of users are sorted as $|h_{1k}| \leq \cdots \leq |h_{N_{\mathrm{RF}}k}|$. The received signal by User (i,k) becomes

$$
y_{ik} = \underbrace{\sqrt{\alpha p_i} h_{ik} s_{ik}}_{\text{intended signal}} + \underbrace{\sum_{l=1,l\neq i}^{N_{\mathrm{RF}}} \sqrt{\alpha p_l} h_{ik} s_{lk}}_{\text{intra-beam interference}} + \underbrace{n_{ik}}_{\text{noise}}, \tag{4.17}
$$

where n_{ik} is the additive white Gaussian noise with variance σ^2. As it is seen, inter-beam interference is removed. In each beam, with using NOMA, successive detection is carried out in descending order. Therefore, the achievable rate of User (i,k) is obtained as

$$
R_{ik} = \log_2\left(1 + \frac{\alpha \rho_i |h_{ik}|^2}{|h_{ik}|^2 \sum_{l=i+1}^{N_{\mathrm{RF}}} \alpha \rho_l + 1}\right). \tag{4.18}
$$

4.6 Power Allocation For Downlink RA-NOMA

In this section, we study power allocation for the proposed RA-NOMA system. Let $\mathbf{p} = [p_1, \ldots, p_{N_{\mathrm{RF}}}]^T$ be the transmission powers of the RAMA groups. Our objective is to optimize the power allocation to maximize the sum achievable rate under the total power and individual minimum rate constraints for users as follows

$$
\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{k=1}^{N_B} \sum_{i=1}^{N_{\mathrm{RF}}} R_{ik} \tag{4.19a}
$$

$$
\text{subject to} \quad \sum_{i=1}^{N_{\mathrm{RF}}} p_i \leq P_{\max}, \tag{4.19b}
$$

$$
R_{ik} \geq \bar{R}_{ik}, \quad \forall i,k, \tag{4.19c}
$$

$$
\mathbf{p} \geq \mathbf{0}, \tag{4.19d}
$$

where, $P_{\max}$ is the maximum transmit power of the system and $\bar{R}_{ik}$ denotes the minimum (required) rate at User (i,k). The defined power allocation problem above can ensure user fairness by assuming the same minimum rate requirement for all users in the constraint (4.19c).

Lemma 3. The maximum of minimum rate requirement in the ith RAMA group defines the minimum power requirement for the ith RF chain.

Proof. See Appendix B.1. □

Using Lemma 3, it is easy to show that each beam achieves the same transmission rate. Therefore, the power optimization problem in (4.19) is transformed into the power allocation over one beam. We define $\bar{R}_i = \max_k \bar{R}_{ik}$, $|h_i| = |h_{i,k}|$, for $k \in \{1,\ldots,N_B\}$, and $a_i = \sum_{l=i}^{N_{\mathrm{RF}}} p_l$. Considering the above, the optimization problem over one beam is as follows.

$$\underset{\mathbf{p}}{\text{maximize}} \quad \sum_{i=1}^{N_{\mathrm{RF}}} \log_2 \left(\frac{|h_i|^2 a_i + \frac{\sigma^2}{\alpha}}{|h_i|^2 a_{i+1} + \frac{\sigma^2}{\alpha}} \right) \tag{4.20a}$$

$$\text{subject to} \quad \sum_{i=1}^{N_{\mathrm{RF}}} p_i \le P_{\max}, \tag{4.20b}$$

$$|h_i|^2 a_i + \sigma^2/\alpha \ge 2^{\bar{R}_i} \left(|h_i|^2 a_{i+1} + \sigma^2/\alpha \right), \forall i, \tag{4.20c}$$

$$\mathbf{p} \ge \mathbf{0}. \tag{4.20d}$$

The optimization problem in (4.20) has a feasible solution if there is any distribution of $\mathbf{p}$ such that satisfies the minimum rate requirements while meeting the constraint in (4.20b). Therefore, we consider minimizing the total transmission power with rate

constraints for users as

$$\underset{\mathbf{p}}{\text{minimize}} \quad \|\mathbf{p}\|_1 \tag{4.21a}$$

$$\text{subject to} \quad R_i \geq \bar{R}_i, \forall i. \tag{4.21b}$$

We define the answer to the optimization problem in (4.21) as $\mathbf{p}^*$ and the feasibility condition of problem (4.20) becomes $\|\mathbf{p}^*\|_1 \leq P_{\max}$. Due to lack of space, we use the results in [99] for solving (4.21) and the optimal power for the p_i^* can be derived as

$$p_i^* = \left(\sum_{l=i+1}^{N_{\text{RF}}} p_l^* + \frac{\sigma^2}{\alpha |h_l|^2} \right) \left(2^{\bar{R}_i} - 1 \right). \tag{4.22}$$

Therefore, the optimal value of (4.21) is $\|\mathbf{p}^*\|_1 = \sum_{i=1}^{N_{\text{RF}}} p_i^*$. By some manipulations we have

$$\|\mathbf{p}^*\|_1 = \sum_{l=1}^{N_{\text{RF}}} \frac{\left(\Pi_{m=1}^{l-1} 2^{\bar{R}_m} \right) \left(2^{\bar{R}_l} - 1 \right) \frac{\sigma^2}{\alpha}}{|h_l|^2} \leq P_{\max}, \tag{4.23}$$

as the feasibility condition of the optimization problem in (4.20).

The objective function in (4.20a) is proved to be concave in [100]. Since the constraints in (4.20) are linear, the optimization problem is convex. Therefore, the solution of the problem can be found by solving the Karush-Kuhn-Tucker (KKT)

conditions [101]. The Lagrangian function of the problem (4.20) is given by

$$
\begin{aligned}
\mathcal{L}\left(\mathbf{p}, \gamma, \boldsymbol{\beta}\right) = & \sum_{i=1}^{N_{\mathrm{RF}}} \log_2 \left(\frac{|h_i|^2 a_i + \frac{\sigma^2}{\alpha}}{|h_i|^2 a_{i+1} + \frac{\sigma^2}{\alpha}} \right) \\
& + \gamma \left(P_{\max} - \sum_{i=1}^{N_{\mathrm{RF}}} p_i \right) \\
& + \sum_{i=1}^{N_{\mathrm{RF}}} \beta_i \left(|h_i|^2 a_i + \frac{\sigma^2}{\alpha} - 2^{\bar{R}_i} \left(|h_i|^2 a_{i+1} + \frac{\sigma^2}{\alpha} \right) \right),
\end{aligned}
\tag{4.24}
$$

where, γ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{N_{\mathrm{RF}}}]^T$ are the non-negative Lagrangian multipliers asso-ciated with (4.20b) and (4.20c), respectively. From this point, the KKT conditions and solution is the same as [100]. Therefore, to save space and avoid replication, we just state the results of the solution as follows

- The Lagrange multiplier γ is greater than zero and constraint (4.20b) holds with equality.

- The Lagrange multipliers β_i for $i = 1, \ldots, N_{\mathrm{RF}-1}$ are greater than zero and the solution of the problem depends on the multiplier $\beta_{N_{\mathrm{RF}}}$.

- If $\beta_{N_{\mathrm{RF}}} > 0$, the constraint (4.20c) holds with equality for $i = 1, \ldots, N_{\mathrm{RF}}$ and the optimal value of problem (4.20) (maximum achievable rate for one beam) is $\sum_{i=1}^{N_{\mathrm{RF}}} \bar{R}_i$ and if $\beta_{N_{\mathrm{RF}}} = 0$, the optimal value is as follows

$$
\log_2 \left(1 + \frac{\alpha P_{\max} |h_{N_{\mathrm{RF}}}|^2}{\sigma^2 \Pi_{i=1}^{N_{\mathrm{RF}}-1} 2^{\bar{R}_i}} - \sum_{i=1}^{N_{\mathrm{RF}}-1} \frac{|h_{N_{\mathrm{RF}}}|^2 \left(2^{\bar{R}_i} - 1 \right)}{|h_i|^2 \Pi_{l=i}^{N_{\mathrm{RF}}-1} 2^{\bar{R}_l}} \right) + \sum_{i=1}^{N_{\mathrm{RF}}-1} \bar{R}_i.
\tag{4.25}
$$

Consequently, the optimal value of problem (4.19) is that of problem (4.20) multiplied by N_B.
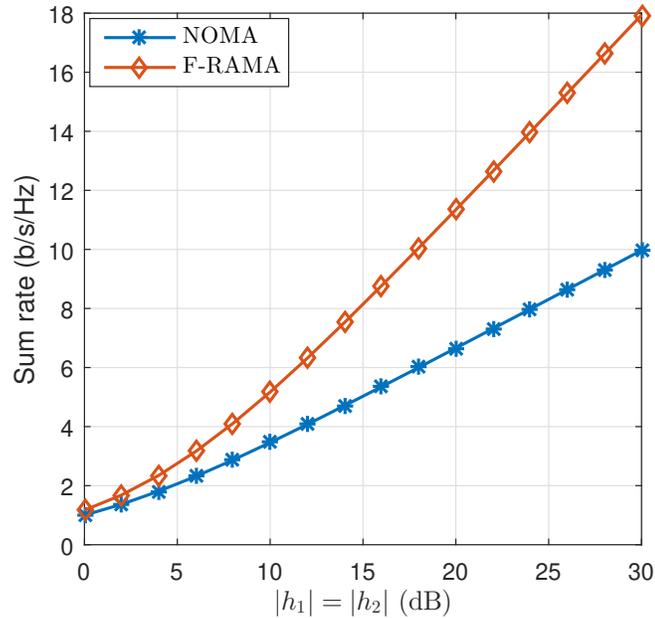
Figure 4.4   sum-rate comparison between NOMA and F-RAMA for symmetric channel.

4.7   Simulation Results

This section evaluates the performance of the proposed multiple access technique by using numerical computations, where the analytical findings will also be verified. Transmission in mmWave bands can be done through both line-of-sight and non line-of-sight paths. Here, for the sake of simplicity, we will consider Rayleigh fading channels.

Figure 4.4 represents sum-rate versus symmetric channel plot for NOMA and F-RAMA. It is clear that F-RAMA achieves better sum-rate than NOMA. This is because the user achievable rate for NOMA is limited due to the interference from other users, which degrades the sum-rate. At symmetric channels, the interference in NOMA is severe and as a result leads to a considerable sum-rate gap compared to
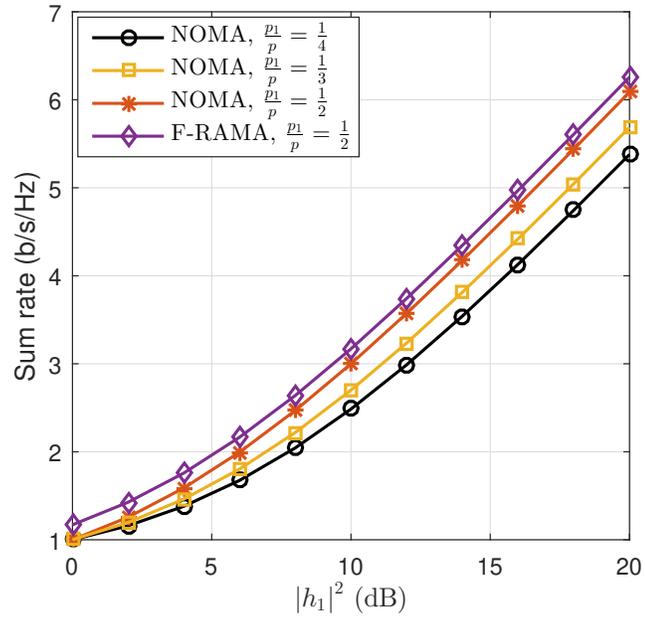
Figure 4.5: sum-rate comparison between NOMA and F-RAMA for asymmetric channel.
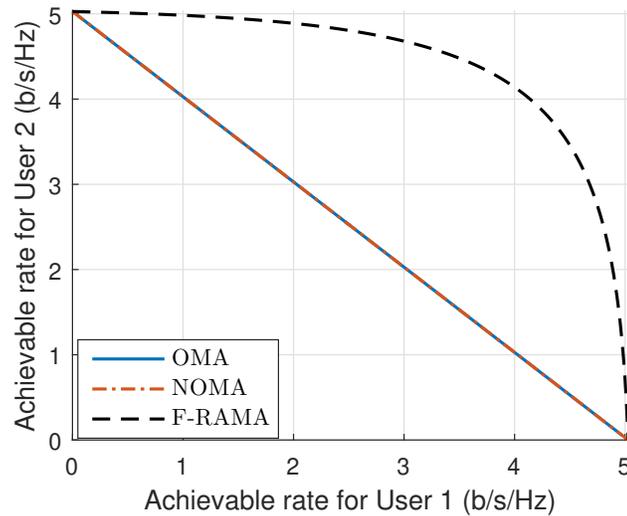
Figure 4.6: Achievable rate region of users 1 and 2 for symmetric channel with $p|h_i|^2/\sigma_i^2 = 15$ dB for $i = 1, 2$.
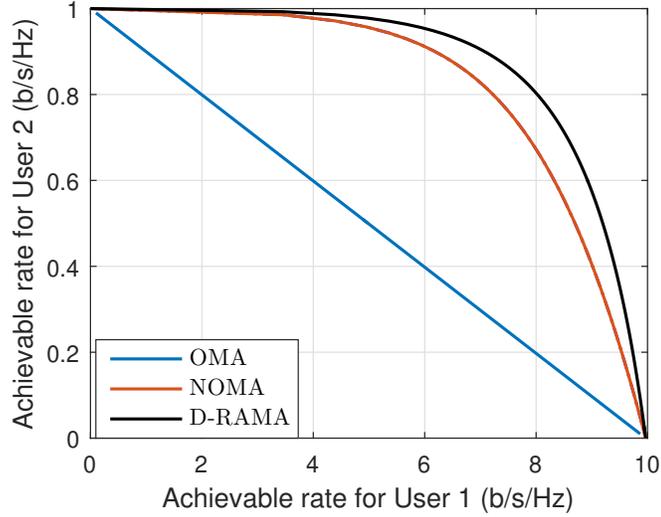
Figure 4.7: Achievable rate region of users 1 and 2 for asymmetric channel with $p|h_1|^2/\sigma^2 = 30$ **dB and** $p|h_2|^2/\sigma^2 = 0$ **dB.**

the RAMA technique which is an inter-user interference-free technique. This result verifies our claim in Case I in Subsection III.A.

Figure 4.5 illustrates sum-rate performance versus asymmetric channel gains for F-RAMA and various power allocation schemes for NOMA. The aim of this simulation is to support our analytical finding in Case II Subsection III.A. When $(p|h_1|^2/\sigma^2)/ (p|h_2|^2/\sigma^2)$ is not large enough, F-RAMA outperforms NOMA for all power allocation schemes since the channel is similar to a symmetric channel. By increasing $(p|h_1|^2/\sigma^2)/(p|h_2|^2/\sigma^2)$ channel satisfies the condition in Case II in Subsection III.A. At high region of $(p|h_1|^2/\sigma^2)/(p|h_2|^2/\sigma^2)$, for $\alpha_2/p \leq 1/2$, e.g., $\alpha_2/p = 1/2$ and $\alpha_2/p = 1/4$, sum-rate of F-RAMA is always better than NOMA which is consistent with Case II. For $\alpha_2/p = 3/4$ and large channel gain difference, NOMA has a little better sum-rate. This is because much more power is allocated to User 1 which can nearly achieve maximum sum-rate. However, in this condition NOMA does not con-

sider user fairness. In contrast, equal power is allocated for the users in F-RAMA and they cannot exploit maximum sum-rate.

Figs. 4.6 and 4.7 show achievable rate region of two users for OMA, NOMA, and D-RAMA. OMA in downlink transmission is assumed to be implemented by OFDMA technique where $R_1^O = \beta \log_2(1 + \alpha_2|h_1|^2/\beta\sigma^2)$ and $R_2^O = (1-\beta)\log_2(1 + \alpha_2|h_2|^2/(1-\beta)\sigma^2)$ are achievable rate for Users 1 and 2 with the bandwidth of β Hz assigned to User 1 and $(1-\beta)$ Hz assigned to User 2 [56].

In Fig. 4.6, channel is assumed to be symmetric where its gain is set to $p|h_i|^2/\sigma^2 = 15$ dB for $i = 1, 2$. The achievable rate region for OMA and NOMA are identical. The region for D-RAMA is much wider than that for OMA and NOMA because D-RAMA neither suffers from inter-user interference nor divides the bandwidth among users. For instance, when User 2 achieves rate 2.5 bits/s/Hz, achievable rate of User 1 for D-RAMA with channel gain information is approximately twice higher than that for OMA and NOMA. Notice that when total power is allocated to either user, all three techniques are able to achieve maximum rate for that user.

The achievable sum-rate of asymmetric channel for $p|h_1|^2/\sigma^2 = 30$ dB and $p|h_2|^2/\sigma^2 = 0$ dB has been represented in Fig. 4.7. From the figure it is clear that for OMA, NOMA and D-RAMA only User 1 achieves maximum sum-rate when whole power is allocated to that user. However, NOMA achieves wider rate region than OMA as expected. Interestingly, the achievable rate region for D-RAMA is greater than NOMA. As an example, when we want User 1 to achieve 8 bits/s/Hz, User 1 can reach 0.68 bits/s/Hz and 0.8 bits/s/Hz with NOMA and D-RAMA, respectively. This is because the achievable rate of User 2 in NOMA is affected by the interference term from User 1 due to using superposition coding at transmitter and SIC process at
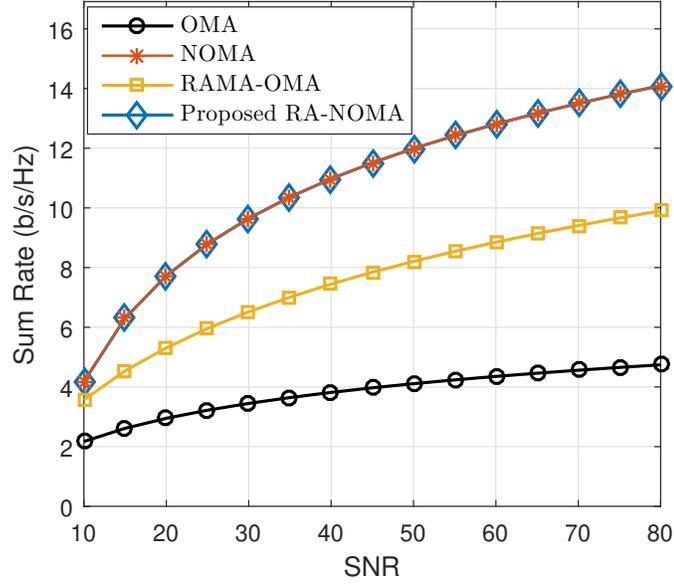
Figure 4.8   Sum-rate comparison for OMA, NOMA, RAMA-OMA, and RA-NOMA with respect to transmit SNR.

Table 4.1   Number of required RF chains and time slots for each technique shown in Fig. 4.8.

	OMA	RAMA-OMA	NOMA	RA-NOMA
Number of RF chains	1	1	4	4
Number of time slots	16	4	1	1

receiver. Whereas, in D-RAMA, User 1 does not impart interference on User 2. In other words, the gap between NOMA and D-RAMA reflects the impact of inter-user interference on NOMA.

In Figs. 4.8 and 4.9, we consider 16 users to be served by various multiple access techniques. The users are scheduled in four RAMA groups and four beams as shown in Fig. 4.3. The minimum rate requirements are considered to be equal and set to $\bar{R}_{i,k} = 0.2$ (b/s/Hz). We define the transmit SNR as the normalized transmit power with respect to σ^2. We assume that the users in RAMA groups have $|h|^2/\sigma^2 =$
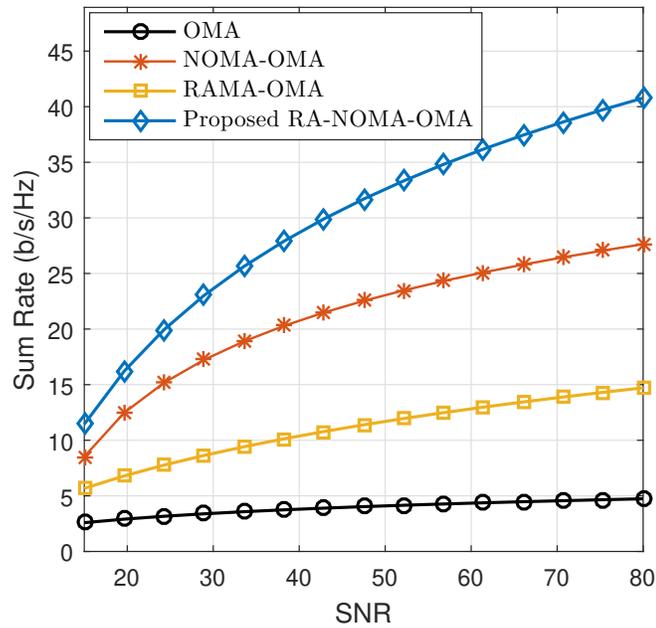
Figure 4.9   Sum-rate comparison for OMA, NOMA-OMA, RAMA-OMA, and RA-NOMA-OMA with respect to transmit SNR.

$\{0, -5, -10, -15\}$ dB.

Fig. 4.8 represents the sum-rate for the four techniques: OMA, NOMA, RAMA-OMA, and the proposed RA-NOMA versus the transmit SNR. Table 4.1 provides the number of required RF chains and time slots for the mentioned techniques. For OMA and RAMA-OMA one RF chain is used, whereas for NOMA and RA-NOMA four RF chains are used. Also, for OMA and RAMA-OMA the transmission is performed in 16 time slots and for NOMA and RA-NOMA the transmission is performed in one time slot. As it is shown, the proposed RA-NOMA and NOMA outperforms OMA and RAMA-OMA techniques in terms of sum-rate with equal power budget at the transmitter. It should be noticed that RA-NOMA has the same sum-rate as that of NOMA. This is because under the considered grouping both users experience the

Table 4.2   Number of required RF chains and time slots for each technique shown in Fig. 4.9.

	OMA	RAMA-OMA	NOMA-OMA	RA-NOMA-OMA
Number of RF chains	1	1	2	2
Number of time slots	16	4	2	2

same amount of power and interference.

Fig. 4.9 represents the sum-rate for the four techniques: OMA, NOMA-OMA, RAMA-OMA, and RA-NOMA-OMA versus the transmit SNR. Table 4.2 provides the number of required RF chains and time slots for the mentioned techniques. For OMA and RAMA-OMA one RF chain is used, whereas for NOMA-OMA and RA-NOMA-OMA two RF chains are used. Also, for OMA and RAMA-OMA the transmission is performed in 16 time slots and for NOMA-OMA and RA-NOMA-OMA the transmission is performed in two time slots. It is revealed that the proposed RA-NOMA-OMA outperforms OMA, RAMA-OMA, and NOMA-OMA. The purpose of this figure is to show that when the number of RF chain is less than the number of beams, the interference is severe for NOMA than that of RA-NOMA. This leads to higher interference in NOMA as the inter-user interference occurs between four users. While, in RA-NOMA, the inter-user interference occurs between two users.

4.8   Summary and Conclusion

In this chapter, we proposed new multiple access techniques for mmWave reconfigurable antennas in order to simultaneously support many users by using a single BS in downlink. First, we show that NOMA is not a suitable technique for serving users for the reconfigurable antenna systems. Then, by wisely using the properties of reconfigurable antennas, a novel multiple access technique called RAMA is designed

by assuming partial CSI and full CSI. The proposed RAMA provides a mmWave reconfigurable antenna system with an inter-user interference-free user serving. That is, the users with higher allocated power are not required to decode signals of other users. It is shown that for symmetric channels F-RAMA outperforms NOMA for an arbitrary α_2 in terms of sum-rate. Also, for asymmetric channels F-RAMA demonstrates better sum-rate performance if approximately more than half of the power is allocated to User 2. Further, D-RAMA always achieves higher sum-rate than NOMA. Further, we proposed RA-NOMA as another multiple access technique. To achieve this, we take advantage of both RAMA and NOMA techniques. Due to the directive and independent beams steered by the lens antennas, the inter-cluster interference between the users is eliminated. The proposed RA-NOMA simultaneously supports a large number of users by using a single BS in the downlink and less number of RF chains compared to the existing mmWave-NOMA technique. The maximum achievable downlink rate of RA-NOMA is derived and the simulation results demonstrate that the proposed RA-NOMA achieves higher sum-rate compared to RAMA and conventional OMA and NOMA techniques.

CHAPTER 5:

IMPACT OF BEAM MISALIGNMENT ON HYBRID BEAMFORMING NOMA FOR MMWAVE COMMUNICATIONS

This chapter analyzes the effect of beam misalignment on rate performance in downlink of hybrid beamforming-based non-orthogonal multiple access (HB-NOMA) systems. First an HB-NOMA framework is designed in multiuser millimeter wave (mmWave) communications. A sum-rate maximization problem is formulated for HB-NOMA, and an algorithm is introduced to design digital and analog precoders and efficient power allocation. Then, regarding perfectly aligned line-of-sight (LoS) channels, a lower bound for the achievable rate is derived. Next, when the users experience misaligned LoS or non-LoS (NLoS) channels, the impact of beam misalignment is evaluated. To this end, a misalignment factor is modeled and each misaligned effective channel is described in terms of the perfectly aligned effective channel parameters and the misalignment factor. Further, a lower bound for the achievable rate is extracted. We then derive an upper bound for the rate gap expression between the aligned and misaligned HB-NOMA systems. The analyses reveal that a large misalignment can remarkably degrade the rate. Extensive numerical simulations are conducted to verify the findings.

5.1   Introduction

Integration of NOMA into mmWave systems, which allows multiple users to share the same beam or the same RF chain, has received considerable research interests [62, 92–94, 102–105]. In [102], a random beamforming technique is designed for mmWave-NOMA systems where the BS randomly radiates a directional beam toward paired users. In [92], it is shown that mismatch between the users' channel vector and finite resolution analog beamforming[1] simplifies utilizing NOMA in MIMO mmWave systems. In [93], a combination of beamspace MIMO and NOMA is proposed to ensure that the number of served users is not limited to the number of RF chains. In [94], NOMA is studied for hybrid mmWave MIMO systems, where a power allocation algorithm has been provided in order to maximize energy efficiency. In all aforementioned works, NOMA is combined with mmWave systems assuming only baseband precoders/combiners. The works in [62, 103–105] have recently studied NOMA in hybrid beamforming systems. Ref. [62] proposes a beam splitting NOMA scheme for hybrid beamforming mmWave systems. In order to increase the spectral efficiency, some users are served with a common RF chain but the grated beams. This technique is only proper when the angle of the directional beams serving the users is large enough. Also, beam grating divides the power of a strong mmWave beam. Hence, far users cannot capture the required power. In [103], designing beamforming vectors and allocating power for just two users have been studied. In [104], it is demonstrated that due to the utilization of HB, the digital precoder of the BS is not perfectly aligned with the user's effective channel. Then, a power

---

[1]Finite resolution analog beamforming is due to the use of a finite number of phase-shifters in the analog beamformer.

allocation algorithm that maximizes the sum-rate has been proposed. Only two users in each beam is considered; moreover, the work fails to study the effect of analog beamforming on the rate performance. Newly, Zhou *et al.* have proposed an angle-based user pairing strategy [105]. The strategy repeatedly switches between NOMA and OMA techniques. Such that, when the beamwidth of the mainlobe of BS is not smaller than the angle difference between two users, they are considered as NOMA users. Otherwise, they are treated as OMA users. Then, the coverage probability and the sum-rate are evaluated. Regularly switching between NOMA and OMA will add more hardware complexity to the system. Also, as it is mentioned, OMA techniques may not be a practical choice for mmWave systems. *In mmWave systems, due to the directional nature of beams in mmWave systems, beam misalignment between the BS and users is inevitable [106].* Most of the reviewed works consider neither the effect of phase-shifters employed in the analog beamformer of a hybrid beamforming system nor the effect of beam misalignment.

5.1.1   Contributions

In this chapter, we investigate the impact of exploiting NOMA in multiuser hybrid beamforming systems termed HB-NOMA. At the outset, it is supposed that HB-NOMA users are paired with respect to their locations and effective channels which is widely adopted by recent research works [62, 92–94, 102–105]. The achievable rate is evaluated when the BS and users' beam are aligned and misaligned. Essentially, the perfect beam alignment is attributed to the existence of LoS channel aligned in the same direction between the BS and users which allows the users to steer their beam directly toward the BS. The imperfect beam alignment (misalignment) occurs due to practical phenomena such as misaligned LoS channels and NLoS channels

which are caused by shadowing and blockage. To the best of authors' knowledge, this study is the first research work that studies the effect of integration of hybrid beamforming and NOMA on the achievable rate in the presence of beam alignment and misalignment. The contribution of this chapter is summarized as follows.

1. We incorporate the 5G enabling technology NOMA and a multiuser hybrid beamforming system studied in [18]. Since we aim to evaluate the impact of beam misalignment on the downlink of HB-NOMA systems, a sum-rate expression is formulated. Specifically, we revise the sum-rate expression in [18] with regard to the NOMA technique. Then, an algorithm is introduced to maximize the system sum-rate subject to a total power constraint, in three steps. To get the first and second steps, we design the analog and digital precoders only regarding LoS channels using the well-known strong effective channel-based effective channel precoder. The third step is a location-based static power allocation.

2. As the maximized sum-rate directly depends on the effective channels of users, we first study the rate for perfect beam alignment where all users exploit LoS channels. A lower bound is derived for the achievable rate of an HB-NOMA user. The bound reveals that the interference is just due to using NOMA in which SC technique at transmitter and SIC at the receiver are exploited. That is to say, the interference on a user is caused by NOMA users located inside the same cluster called intra-cluster interference. Indeed, hybrid beamforming slightly amplifies the noise term which is led by analog devices used in the beamformer. The analysis shows that for the perfect alignment, the HB-NOMA users can achieve a rate which is close to that of NOMA with the fully-digital

beamforming systems.

3. We study the achievable rate of the maximized sum-rate for misaligned beams between the BS and users in the presence of misaligned LoS and NLoS channels. Toward this goal, the beam misalignment problem is modeled by a beam misalignment factor. Considering the derived factor, the effective channel of the users with misaligned LoS or NLoS channel is described in terms of the aligned effective channel parameter and the misalignment factor.

4. We extract a lower bound for the achievable rate using the effective channel model. Three terms, i.e., intra-cluster interference, inter-cluster interference, and noise, constrain the achievable rate. Unfortunately, these terms are directly or indirectly associated with misalignment factors. It is concluded that in HB-NOMA with the precoder based on the strongest effective channel the achievable rate of a user depends on both the effective channel gain and beam alignment issue. This is opposite to the fully-digital NOMA systems in which only the effective channel gain affects the rate. Then, an upper bound for rate gap between the aligned and misaligned HB-NOMA user is found.

To confirm the analyses and the derived expressions, numerical simulations are done. Different HB-NOMA system parameters are evaluated. The simulations indicate that the HB-NOMA outperforms OMA.

The chapter is organized as follows: Section 5.2 presents the system model of HB-NOMA and formulates a sum-rate expression. In Section 5.3, we maximize the sum-rate for perfect beam alignment then analyze the rate performance. Section 5.4 studies the rate performance for beam misaligned HB-NOMA. In Section 5.5, we present

simulation results investigating the rate performance of HB-NOMA. Section 5.6 concludes this chapter.

Notations: Hereafter, $j = \sqrt{-1}$, small letters, bold letters and bold capital letters will designate scalars, vectors, and matrices, respectively. Superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^\dagger$ denote the transpose, conjugate and transpose-conjugate operators, respectively. Further, $|\cdot|$, $\|\cdot\|$, and $\|\cdot\|_2$ denote the absolute value, norm-1 of $(\cdot)$, and norm-2 of vector $(\cdot)$, respectively. Indeed, $\|\cdot\|_F$ denotes the Frobenius norm of matrix $(\cdot)$. Finally, $\mathbb{E}[\cdot]$ denotes the expected value of $(\cdot)$.

5.2 System Model and Rate Formulation

5.2.1 System Model for HB-NOMA

We assume a narrow band mmWave downlink system composed of a BS and multiple users as shown in Fig. 5.1. The BS is equipped with N_{RF} chains and N_{BS} antennas whereas each user has one RF chain and N_{UE} antennas. Each RF chain is connected to the antennas through phase-shifters. We also assume that the BS communicates with each user via only one stream. This will be justified later in the present section. In traditional multiuser systems based on the hybrid beamforming the maximum number of users that can be simultaneously served by the BS equals the number of BS RF chains [18].

In order to establish a better connectivity in dense areas and further improve the sum-rate, this chapter develops the HB-NOMA system. The system is practical and takes the parameters of the promising hybrid beamforming into account. To achieve this, we utilize NOMA in hybrid beamforming multiuser systems where each beam is allowed to serve more than one user. The transmitter simultaneously sends
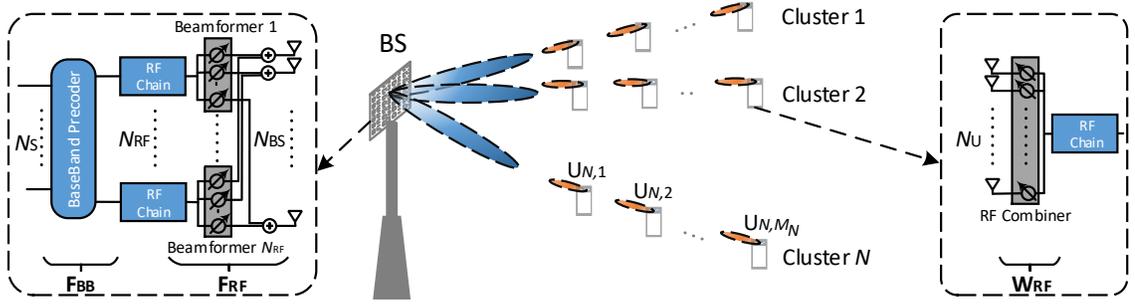
Figure 5.1   **HB-NOMA with one BS and huge number of users grouped into N clusters each with M_n NOMA users. $N_{\mathbf{S}}, N_{\mathbf{RF}}, N_{\mathbf{BS}}$, and $N_{\mathbf{UE}}$ are the numbers of multiplexed streams, RF chains, BS antennas, and user antennas, respectively.**

N_{S} streams toward $\sum_{n=1}^{N} M_n$ users which are grouped into $N \leq N_{\mathrm{RF}}$ clusters. M_n denotes the number of users in the nth cluster. The users in each cluster can be scheduled by using the efficient approaches presented in [107, 108]. Without loss of generality, we assume $N_{\mathrm{S}} = N$. Hence, $\sum_{n=1}^{N} M_n \gg N_{\mathrm{RF}}$; i.e., an HB-NOMA system can simultaneously serve $\sum_{n=1}^{N} M_n$ users which is much larger than the number of RF chains. In the following we formulate the transmit and received signals for the HB-NOMA system.

Superposition coding

On the downlink of the HB-NOMA system, first, the transmit symbols are superposition coded at the BS. Let $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$ denote the information signal vector such that $\mathbb{E}\left[s_n s_n^*\right] = \frac{1}{N}$. Each $s_n = \sum_{m=1}^{M_n} \sqrt{P_{n,m}} s_{n,m}$ is the superposition coded signal performed by NOMA with $P_{n,m}$ and $s_{n,m}$ being transmit power and transmit information signal for the mth user in the nth cluster. Then, the hybrid beamforming is done in two stages. In the first stage, the transmitter applies an $N \times N$ baseband

precoder $\mathbf{F}_{\mathrm{BB}}$ using its N_{RF} RF chains. This stage then is followed by an $N_{\mathrm{BS}} \times N$ RF precoder $\mathbf{F}_{\mathrm{RF}}$ using analog phase-shifters. Thus, the transmit signal vector after superposition coding is given by

$$[x_1, x_2, \ldots, x_N]^T = \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} [s_1, s_2, \ldots, s_N]^T, \tag{5.1}$$

where x_n denotes the transmit signal toward the nth cluster. Hereafter, $\mathrm{UE}_{n,m}$ denotes the mth user in the nth cluster. Since $\mathbf{F}_{\mathrm{RF}}$ is implemented by using analog phase-shifters it is assumed that all elements of $\mathbf{F}_{\mathrm{RF}}$ have an equal norm, i.e., $|(\mathbf{F}_{\mathrm{RF}})_{n,m}|^2 = N_{\mathrm{BS}}^{-1}$. Also, the total power of the hybrid transmitter is limited to $\left\|\mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}}\right\|_F^2 = N$ [18, 33].

Successive interference cancellation

The received signal at $\mathrm{UE}_{n,m}$ is given by

$$\mathbf{r}_{n,m} = \mathbf{H}_{n,m} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \mathbf{s} + \mathbf{n}_{n,m}, \tag{5.2}$$

where $\mathbf{H}_{n,m}$ of size $N_{\mathrm{UE}} \times N_{\mathrm{BS}}$ denotes the mmWave channel between the BS and $\mathrm{UE}_{n,m}$ such that $\mathbb{E}\left[\left\|\mathbf{H}_{n,m}\right\|_F^2\right] = N_{\mathrm{BS}} N_{\mathrm{UE}}$. The channel matrix $\mathbf{H}_{n,m}$ is described in (2.1) where only the strongest NLoS channel is considered. That is, $\mathbf{H}_{n,m} = \beta_{n,m} \mathbf{a}_{\mathrm{UE}}(\vartheta_{n,m}) \mathbf{a}_{\mathrm{BS}}^\dagger(\varphi_{n,m})$, where $\beta_{n,m} = g_{n,m} d_{n,m}^{\frac{-\nu}{2}}$ with $g_{n,m}$ is the complex gain with zero-mean and unit-variance. Further, the antenna configuration is assumed to be uniform linear array (ULA). Thus, only azimuth angles information are taken into account. $\mathbf{n}_{n,m} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ is the additive white Gaussian noise vector of size $N_{\mathrm{UE}} \times 1$. Each component of $\mathbf{n}_{n,m}$ has zero-mean and σ^2 variance. $\mathbf{I}$ denotes the

identity matrix of size $N_{\text{UE}} \times N_{\text{UE}}$. At $\text{UE}_{n,m}$, the RF combiner is used to process the received vector as

$$
y_{n,m} = \underbrace{\mathbf{w}_{n,m}^{\dagger}\mathbf{H}_{n,m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}^{n}\sqrt{P_{n,m}}s_{n,m}}_{\text{desired signal}} + \underbrace{\mathbf{w}_{n,m}^{\dagger}\mathbf{H}_{n,m}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}^{n}\sum_{k=1,k\neq m}^{M}\sqrt{P_{n,k}}s_{n,k}}_{\text{intra-cluster interference}}
$$
$$
+ \underbrace{\mathbf{w}_{n,m}^{\dagger}\mathbf{H}_{n,m}\sum_{\ell=1,\ell\neq n}^{N}\mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}^{\ell}\sum_{q=1}^{M}\sqrt{P_{\ell,q}}s_{\ell,q}}_{\text{inter-cluster interference}} + \underbrace{\mathbf{w}_{n,m}^{\dagger}\mathbf{n}_{n,m}}_{\text{noise}}, \tag{5.3}
$$

where $\mathbf{w}_{n,m} \in \mathbb{C}^{N_{\text{UE}}\times 1}$ denotes the combiner at $\text{UE}_{n,m}$. After combining, each user decodes the intended signal by using SIC as follows. The first user of each cluster, which has the highest channel gain, is allocated the lowest power and the M_nth user, which has the lowest channel gain, is allocated the highest power. At the receiver side, $\text{UE}_{n,m}$ decodes the intended signal of $\text{UE}_{n,k'}$, i.e., $s_{n,k'}$, for $k' = m+1, m+2, \ldots, M_n$ and subtracts it from the received signal $y_{n,m}$. However, NOMA treats the intended signal of $\text{UE}_{n,k}$ for $k = 1, 2, \ldots, m-1$ as intra-cluster interference. In this chapter, SIC process is assumed to be ideal. When SIC is non-ideal, the user cannot completely remove the signals of some of $\text{UE}_{n,k'}$ for $k' = m_1, m+2, \ldots, M_n$ which degrades the performance of the system [109]. The effect of non-ideal SIC on NOMA has recently been studied in [110]. The effect of non-ideal SIC on HB-NOMA will be evaluated in the authors' future work. To this end, the BS should send the order of superposition coding to all users in the cluster. Usually NOMA users are selected to have very different channel gains, specially in mmWave frequencies in which path loss is higher that sub-6 GHz frequencies. So, the order of decoding can be estimated from the user's distance to the BS or its channel gain, correspondingly. We note that the order

of encoding is related to the channel gain as indicated in Section 5.3.1.

5.2.2 Rate Formulation

In (5.3), after applying superposition coding at the transmitter, each user experiences two types of interference. Intra-cluster interference which is due to other users within the cluster and inter-cluster interference which is due to users within other clusters. Suppressing the intra-cluster interference directly depends on efficient power allocation and deploying SIC which is discussed in the previous section. To mitigate the inter-cluster interference, the transmitter needs to design a proper beamforming matrix which will be discussed in Section 5.3.1. Hence, the rate for $\text{UE}_{n,m}$ is expressed as

$$R_{n,m} = \log_2 \left(1 + \frac{P_{n,m} \left| \mathbf{w}_{n,m}^{\dagger} \mathbf{H}_{n,m} \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}}^{n} \right|^2}{I_{\text{intra}}^{n,m} + I_{\text{inter}}^{n,m} + \sigma^2} \right), \tag{5.4}$$

where $I_{\text{intra}}^{n,m}$ is given by

$$I_{\text{intra}}^{n,m} = \sum_{k=1}^{m-1} P_{n,k} \left| \mathbf{w}_{n,m}^{\dagger} \mathbf{H}_{n,m} \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}}^{n} \right|^2, \tag{5.5}$$

denotes the intra-cluster. Also, $I_{\text{inter}}^{n,m}$ is defined as

$$I_{\text{inter}}^{n,m} = \sum_{\ell=1,\ell \neq n}^{N} \sum_{q=1}^{M_n} P_{\ell,q} \left| \mathbf{w}_{n,m}^{\dagger} \mathbf{H}_{n,m} \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB}}^{\ell} \right|^2, \tag{5.6}$$

denotes the inter-cluster interference.

5.3 Perfect Beam Alignment: Rate Maximization and Analysis

5.3.1 The Maximization Algorithm

To optimize the sum-rate performance, hybrid precoder $\mathbf{F}_{\mathrm{RF}}$, and $\mathbf{F}_{\mathrm{BB}}$, combiner $\mathbf{w}_{n,m}$ and transmit power $P_{n,m}$ for $m = 1, 2, \ldots, M_n$ and $n = 1, 2, \ldots, N$ should be found from

$$\underset{\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{w}_{n,m}, P_{n,m}}{\text{maximize}} \quad \sum_{n=1}^{N} \sum_{m=1}^{M_n} R_{n,m} \tag{5.7a}$$

$$\text{subject to} \quad \left| (\mathbf{F}_{\mathrm{RF}})_{n,m} \right|^2 = N_{\mathrm{BS}}^{-1}, \tag{5.7b}$$

$$\left\| \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{BB}} \right\|_F^2 = N, \tag{5.7c}$$

$$\left| \mathbf{w}_{n,m} \right|^2 = N_{\mathrm{UE}}^{-1}, \tag{5.7d}$$

$$R_{n,m} \geq R_{\min}, \tag{5.7e}$$

$$\sum_{n=1}^{N} \sum_{m=1}^{M_n} P_{n,m} \leq P, \tag{5.7f}$$

$$P_{n,m} > 0, \tag{5.7g}$$

where P equals to the total transmit power and $R_{\min}$ denotes the predefined minimum rate. In the above optimization problem, the constraints (5.7b) and (5.7d) ensure that all elements of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{w}_n$ have an equal norm. Further, the constraint (5.7c) ensures that the total power of the hybrid transmitter is limited to N. The constraint (5.7e) ensures that the minimum achievable rate of each user is equal or greater than $R_{\min}$. Finally, the constraint (5.7f) guarantees that the total transmit power is limited to

P and the constraint (5.7g) guarantees that the allocated power to $\text{UE}_{n,m}$ is greater than zero. One would add fairness constraint to the maximization problem. Ref. [110] discusses a viable solution in this case. In particular, a weighted sum-rate which considers a special priority for each user is utilized. Also, without of loss of generality, here, we assume that all the users satisfy $R_{n,m} \geq R_{\min}$. In this case, an iterative algorithm that properly allocates the power is required [111].

It is mentioned that transmission in mmWave bands happens through LoS and NLoS channels. In particular, the users which are located far from the BS will mostly be supported via NLoS channels [63]. Let's first focus on only LoS channels. We assume that all channels are LoS and the effective channels are perfectly aligned as shown in Fig. 5.1. By perfect alignment we mean that $\mathbf{a}_{\text{BS}}(\varphi_{n,m})$ is identical for all users in the nth cluster, i.e., $\mathbf{a}_{\text{BS}}(\varphi_{n,1}) = \mathbf{a}_{\text{BS}}(\varphi_{n,2}) = \cdots = \mathbf{a}_{\text{BS}}(\varphi_{n,M_n})$ for $n = 1, 2, \ldots, N$.

In general, there are two extreme cases to design baseband precoder for mmWave-NOMA systems, strong effective channel-based and singular value decomposition (SVD)-based precoder methods [93]. The strong effective channel-based is designed for only LoS channels and the SVD-based precoder is designed for only NLoS channels. Further, to the best of authors' knowledge, it is not shown how to design the SVD-based RF precoder for hybrid beamforming system. Here, in order to understand the behavior of beam misalignment in HB-NOMA systems we choose the strong effective channel-based precoder which is widely used in the literature [93, 94, 104].

The maximization problem in (5.7) is non-convex and finding the optimal solution is not trivial. To ease, we present an efficient and simple algorithm in three steps as described below.

In the first step, the BS and $\text{UE}_{n,m}$ solve the following problem

$$\underset{\mathbf{w}_{n,m},\mathbf{f}_{\text{RF}}^{n,m}}{\text{maximize}} \quad \left| \mathbf{w}_{n,m}^{\dagger} \mathbf{H}_{n,m} \mathbf{f}_{\text{RF}}^{n,m} \right| \qquad \text{subject to (5.7b) and (5.7d)}. \tag{5.8}$$

Since the channel $\mathbf{H}_{n,m}$ has only one path, and given the continuous beamsteering capability assumption $\mathbf{w}_{n,m} = \mathbf{a}_{\text{UE}}(\vartheta_{n,m})$ and $\mathbf{f}_{\text{RF}}^{n,m} = \mathbf{a}_{\text{BS}}(\varphi_{n,m})$, are the optimal solutions [18]. We design the RF (analog) and baseband (digital) precoders using the adopted strong effective channel-based method. Hence, in order to design the RF precoder, the BS selects the first user of each cluster. The RF precoder of the first user of the nth cluster makes the nth column of the RF precoding matrix, i.e., $\mathbf{f}_{\text{RF}}^{n,1}$, gives the RF precoding matrix as

$$\mathbf{F}_{\text{RF}} = \left[\mathbf{f}_{\text{RF}}^{1,1}, \mathbf{f}_{\text{RF}}^{2,1}, \dots, \mathbf{f}_{\text{RF}}^{N,1} \right]. \tag{5.9}$$

The first user is determined based on the locations of the user as follows:

$$|\beta_{n,1}| \geq |\beta_{n,2}| \geq \cdots \geq |\beta_{n,M_n}|, \quad \text{for} \quad n = 1, 2, \dots, N, \tag{5.10}$$

where $\beta_{n,m}$ is the channel fading coefficient. To determine the first user, the BS does not need to know the channel gain of the users. Recall that the channel fading coefficient $\beta_{n,m}$, mainly depends on distance between the BS and $\text{UE}_{n,m}$ (d) and path loss factor (ν). Since the path loss factor is identical for all users, the first user of each cluster can be determined as the closest user to the BS such that its channel gain has the highest amplitude among the users in the same cluster. While the purpose of ordering in (5.10) is to define the first user, to realize NOMA, another ordering

method based on the effective channel gain is presented in the third step. It should be stressed that the main reason to design the digital precoder with respect to the strongest channel is that the strongest user must decode the other users' signal before its signal. So, the power of this user's signal is not affected by other clusters' signal. More details will be provided in Section 5.4.

In the second step, the effective channel for $\text{UE}_{n,m}$ is expressed as

$$\bar{\mathbf{h}}_{n,m}^{\dagger} = \mathbf{w}_{n,m}^{\dagger} \mathbf{H}_{n,m} \mathbf{F}_{\text{RF}} = \beta_{n,m} \mathbf{a}_{\text{BS}}^{\dagger}(\varphi_{n,m}) \mathbf{F}_{\text{RF}}. \tag{5.11}$$

Regarding the strongest channel-based method, we write the effective channel matrix as

$$\overline{\mathbf{H}} = \left[\bar{\mathbf{h}}_{1,1}, \bar{\mathbf{h}}_{2,1}, \ldots, \bar{\mathbf{h}}_{N,1} \right]^{\dagger}, \tag{5.12}$$

where $\bar{\mathbf{h}}_{n,1}$ denotes the effective channel vector of $\text{UE}_{n,1}$.

Designing a proper digital precoder $\mathbf{F}_{\text{BB}}$ can reduce the inter-cluster interference. In brief, designing the baseband precoder becomes equivalent to solving

$$\underset{\{\mathbf{f}_{\text{BB}}^{\ell}\}_{\ell \neq n}}{\text{minimize}} \, I_{\text{inter}}^{n,m} \qquad \text{subject to (5.7c).} \tag{5.13}$$

where $I_{\text{inter}}^{n,m}$ is defined in (5.5). We notice that so far we have designed the analog beamformer and combiner. The only unknown parameter is the digital beamformer. In this chapter, we adopt zero-forcing beamforming (ZFBF) which makes a balance between implementation complexity and performance [112,113]. Based on ZFBF, the solution for (5.13) is obtained as [18]

$$\mathbf{F}_{\text{BB}} = \overline{\mathbf{H}}^{\dagger} \left(\overline{\mathbf{H}} \overline{\mathbf{H}}^{\dagger} \right)^{-1} \mathbf{\Gamma}, \tag{5.14}$$

where the diagonal elements of $\mathbf{\Gamma}$ are given by [18]

$$\mathbf{\Gamma}_{n,n} = \sqrt{\frac{1}{(\mathbf{F}^{-1})_{n,n}}} \, |\beta_{n,1}|, \quad \text{for} \quad n = 1, 2, \ldots, N. \tag{5.15}$$

where $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}^{\dagger} \mathbf{F}_{\mathrm{RF}}$. The determined precoder in (5.14) indicates that inter-cluster interference on first users is zero, i.e., $\bar{\mathbf{h}}_{n,1}^{\dagger} \mathbf{f}_{\mathrm{BB}}^{\ell} = 0$ for $n = 1, 2, \ldots, N$ and $\ell \neq n$. That is, inter-cluster interference is perfectly eliminated on the first users. This completes our justification about the orienting the beams toward the first users and choosing their effective channel vector in designing $\mathbf{F}_{\mathrm{BB}}$.

In the third step, the BS first reorders the users then allocates the power. The reordering process is done based on the effective channel vectors as

$$\left\| \bar{\mathbf{h}}_{n,1} \right\| \geq \left\| \bar{\mathbf{h}}_{n,2} \right\| \geq \cdots \geq \left\| \bar{\mathbf{h}}_{n,M_n} \right\|, \quad \text{for} \quad n = 1, 2, \ldots, N. \tag{5.16}$$

Notice that in (5.10) we aimed to find the first users based on the large-scale gain. However, in HB-NOMA the power allocation is conducted based on order of the effective channel gains. It is not irrational to assume that the BS knows the effective channels. This can be done through the channel quality indicator (CQI) messages [114]. Each user feeds the effective channel back to the BS then it sorts the users.

The optimal power allocation in (5.7) can be done by solving the following problem.

$$\underset{P_{n,m}}{\text{maximize}} \sum_{n=1}^{N} \sum_{m=1}^{M_n} R_{n,m} \qquad \text{subject to (5.7f) and (5.7g).} \tag{5.17}$$

To solve the problem, we propose a two-stage solution. First the BS divides the power

between the clusters considering their users' channel gain as follows.

$$P_n = \frac{\sum_{m=1}^{M_n} \left\| \bar{\mathbf{h}}_{n,m} \right\|^2}{\sum_{n=1}^{N} \sum_{m=1}^{M_n} \left\| \bar{\mathbf{h}}_{n,m} \right\|^2} P, \quad \text{for} \quad n = 1, 2, \ldots, N. \tag{5.18}$$

Then a fixed power allocation is utilized for the users in each cluster respecting the constraint $\sum_{m=1}^{M_n} P_{n,m} = P_n$. To determine $P_{n,m}$, one solution is to allocate a certain amount of power for each $\text{UE}_{n,m}$ except the first one that only satisfies $R_{n,m} = R_{\min}$, then the remaining is assigned to $\text{UE}_{n,1}$. This power allocation process is in consist with the concept of NOMA in which, to achieve higher sum-rate, the stronger user should receive more power [53–56]. On the other hand, recall that mmWave channels are vulnerable to blockage and shadowing. Especially, for the weak users which are located far from the BS, this issue becomes worse. So, the weak users may not be able to achieve the required minimum rate. Another solution is to give priority to the fairness issue. To this, we need to allocate less power to the strong users and more power to the weak users. It turns out, fairness works against achieving maximum rate. Thus, our solution to achieve maximum rate and compensate for the mmWave propagation issues is to assign the same amount of power for all the users, i.e.,

$$P_{n,1} = P_{n,2} = \cdots = P_{n,M_n}. \tag{5.19}$$

5.3.2 The Achievable Rate Analysis

In this section, the achievable rate of $\text{UE}_{n,m}$ is evaluated with respect to the designed parameters. We derive a lower bound which characterizes insightful results on the

achievable rate of HB-NOMA.

Theorem 4. With perfect beam alignment, a lower bound on the achievable rate of $\text{UE}_{n,m}$ is given by

$$
\overline{R}_{n,m} \geq \log_2 \left(1 + \frac{P_{n,m} \left| \beta_{n,m} \right|^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k} \left| \beta_{n,m} \right|^2 + \sigma^2 \kappa_{\min}^{-1}(\mathbf{F})} \right), \tag{5.20}
$$

$\kappa_{\min}(\mathbf{F})$ denotes the minimum eigenvalue of $\mathbf{F}$.

Proof. Please see Appendix C.1. □

Remark 11. Theorem 1 indicates that when the alignment between the users in each cluster is perfect, still two terms degrade the sum-rate performance of every HB-NOMA user. The first term $\sum_{k=1}^{m-1} P_{n,k} \left| \beta_{n,m} \right|^2$ is due to using NOMA scheme which leads to inevitable intra-cluster interference. The second term $\kappa_{\min}^{-1}(\mathbf{F})$ is due to realizing the beamforming with digital and analog components, i.e., hybrid beamforming instead of fully-digital components. It is worth mentioning that in the fully-digital beamforming the first term exists but the second term is always one. Therefore, even under perfect beam alignment assumption the hybrid beamforming intrinsically imposes small loss on the achievable rate.

5.4 Beam Misalignment: Modeling, Rate Analysis, and Rate Gap

In the previous section we designed the precoders when only LoS channels exist and the users are perfectly aligned. The precoders are found based on the strongest
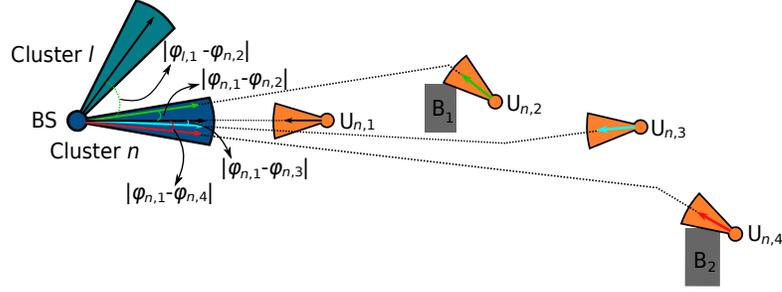
Figure 5.2   **Beam misalignment in mmWave communications due to the NLoS channels. The NLoS channels are caused by blockages B1 and B2.**

effective channel. Perfect alignment is an ideal assumption. In fact, AoDs/AoAs are random variable and with almost surely the probability of occurring different AoDs/AoAs even in LoS channels is one which leads to $\mathbf{a}_{\mathrm{BS}}(\varphi_{n,1}) \neq \mathbf{a}_{\mathrm{BS}}(\varphi_{n,2}) \neq \ldots \neq \mathbf{a}_{\mathrm{BS}}(\varphi_{n,M_n})$ for $n = 1, 2, \ldots, N$. On the other hand, recall that in mmWave frequencies, due to shadowing and blockage, NLoS channels are inevitable [63]. These channels force the users to indirectly steer their beam toward the BS as illustrated by Fig. 5.2. So, the misalignment between the effective channel of the first user and the users with misaligned LoS and NLoS channel in each cluster causes the digital baseband precoder cannot eliminate the inter-cluster interference. As a result, the achievable rate is degraded. In this section, first the misalignment is modeled. Second, using the derived model, a lower bound is found for the rate. Finally, an upper bound is extracted for the rate gap between the perfect alignment and misalignment.

Remark 12. While our findings in this section are general and hold for misaligned LoS and NLoS channels, we only concentrate on NLoS channels. Thus, by LoS channel we mean a perfectly aligned channel. Also, it is assumed that all users expect the first one in all clusters have NLoS channels. In order to distinguish effective channel of the users with aligned LoS channels from NLoS channels, hereafter, we denote $\bar{\mathbf{h}}_{n,m}$

as effective channel of the user with perfect beam alignment and $\tilde{\mathbf{h}}_{n,m}$ as effective channel of the user with imperfect beam alignment. Also, $\overline{R}_{n,m}$ and $\tilde{R}_{n,m}$ denote the rate of $\mathrm{UE}_{n,m}$ with LoS and NLoS channel, respectively.

5.4.1 Beam Misalignment Modeling

In what follows, we study the impact of imperfect beam alignment on the rate. Before that, we calculate the norm of the effective channel defined in (5.11). Defining

$$\left|\mathbf{a}_{\mathrm{BS}}^{\dagger}(\varphi_{n,m})\mathbf{a}_{\mathrm{BS}}(\varphi_{\ell,1})\right|^{2} = K_{N_{\mathrm{BS}}}(\varphi_{\ell,1} - \varphi_{n,m}), \tag{5.21}$$

where $K_{N_{\mathrm{BS}}}$ is Fejér kernel of order N_{BS} [115], we get

$$\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2} = |\beta_{n,m}|^{2} \sum_{\ell=1}^{N} K_{N_{\mathrm{BS}}} \left(\varphi_{\ell,1} - \varphi_{n,m}\right). \tag{5.22}$$

Now, we model the correlation between the effective channels for $\mathrm{UE}_{n,m}$ and $\mathrm{UE}_{n,1}$ and between $\mathrm{UE}_{n,m}$ and $\mathrm{UE}_{\ell,1}$ with $\ell \neq n$ by defining them as intra-cluster misalignment factor and inter-cluster misalignment factor, respectively. Notice that we consider the worst scenario. That is, $\mathrm{UE}_{n,m}$ for $m = 2, 3, \ldots, M_n$ receives the signal through NLoS channel, while only $\mathrm{UE}_{n,1}$ for $n = 1, 2, \ldots, N$ receives through LoS channel. Assuming LoS channel for the first users is reasonable, since in mmWave communications the users close to the BS experience LoS channels with high probability [63].

Lemma 4. The misalignment effective channel of $\mathrm{UE}_{n,m}$ and $\mathrm{UE}_{n,1}$ can be modeled as

$$\hat{\tilde{\mathbf{h}}}_{n,m} = \rho_{n,m}\hat{\tilde{\mathbf{h}}}_{n,1} + \sqrt{1 - \rho_{n,m}^{2}}\,\hat{\mathbf{g}}_{\mathrm{BS}}^{-n}, \tag{5.23}$$

where $\hat{\tilde{\mathbf{h}}}_{n,m}$ denotes the normalized imperfect effective channel, $\rho_{n,m}$ denotes the mis-

alignment factor obtained as

$$\rho_{n,m} = \frac{\sum_{i=1}^{N} \kappa_i(\mathbf{F}) \left| \mathbf{a}_{\mathrm{BS}}^\dagger(\varphi_{n,m}) \mathbf{v}_1^i \mathbf{v}_1^{i\dagger} \mathbf{a}_{\mathrm{BS}}(\varphi_{n,1}) \right|}{\sqrt{\sum_{\ell=1}^{N} K_{N_{\mathrm{BS}}}(\varphi_{\ell,1} - \varphi_{n,m})} \sqrt{\sum_{\ell=1}^{N} K_{N_{\mathrm{BS}}}(\varphi_{\ell,1} - \varphi_{n,1})}}, \tag{5.24}$$

where $\kappa_i(\mathbf{F})$ is the ith eigenvalue of $\mathbf{F}$. $\hat{\mathbf{g}}_{\mathrm{BS}}^{-n}$ is a normalized vector located in the subspace generated by linear combination of $\mathbf{a}_{\mathrm{BS}}(\varphi_{\ell,1})$ for $\ell \neq n$, such that $\hat{\mathbf{g}}_{\mathrm{BS}}^{-n} = \frac{\mathbf{g}_{\mathrm{BS}}^{-n}}{\left\| \mathbf{g}_{\mathrm{BS}}^{-n} \right\|}$, where $\mathbf{g}_{\mathrm{BS}}^{-n} = \mathbf{F}_{\mathrm{RF}}^\dagger \sum_{\ell=1, \ell \neq n}^{N} \beta_{\ell,1} \mathbf{a}_{\mathrm{BS}}(\varphi_{\ell,1})$.

Proof. Please see Appendix C.2. □

5.4.2 Rate Analysis

Now we are ready to find a lower bound for the achievable rate of $\mathrm{UE}_{n,m}$.

Theorem 5. With imperfect beam alignment, a lower bound on the achievable rate of $\mathrm{UE}_{n,m}$, is given by

$$\tilde{R}_{n,m} \geq \log_2 \left(1 + \frac{P_{n,m} \rho_{n,m}^2 |\beta_{n,m}|^2}{\zeta_{\mathrm{intra}}^{n,m} + \zeta_{\mathrm{inter}}^{n,m} + \zeta_{\mathrm{noise}}^{n,m}} \right), \tag{5.25}$$

where $\zeta_{\mathrm{intra}}^{n,m} = \sum_{k=1}^{m-1} P_{n,k} \rho_{n,m}^2 |\beta_{n,m}|^2$ and $\zeta_{\mathrm{inter}}^{n,m} = \left(1 - \rho_{n,m}^2\right) |\beta_{n,m}|^2 \kappa_{\max}(\mathbf{S}) \kappa_{\min}^{-1}(\mathbf{F}) \times K_{N_{\mathrm{BS}},1}$ in which $\kappa_{\max}(\mathbf{S})$ is the maximum eigenvalue of $\mathbf{S} = \mathbf{F}_{\mathrm{BB}}^{-n,W} \mathbf{F}_{\mathrm{BB}}^{-n,W\dagger}$, $\mathbf{F}_{\mathrm{BB}}^{-n,W}$ denotes the wieghted $\mathbf{F}_{\mathrm{BB}}$ after eliminating the nth column where the columns are scaled by $P_\ell \; \forall \ell \neq n$. Also, for some m we define

$$K_{N_{\mathrm{BS}},m} = \sum_{\ell=1}^{N} K_{N_{\mathrm{BS}}}(\varphi_{\ell,1} - \varphi_{n,m}), \tag{5.26}$$

where $K_{N_{\mathrm{BS}}} \left(\varphi_{\ell,1} - \varphi_{n,m} \right)$ denotes the Fejér kernel in (5.21). Finally, $\zeta_{\mathrm{noise}}^{n,m}$ is expressed as $\zeta_{\mathrm{noise}}^{n,m} = \sigma^2 \kappa_{\mathrm{min}}^{-1}(\mathbf{F}) K_{N_{\mathrm{BS}},1} K_{N_{\mathrm{BS}},m}^{-1}$, where $K_{N_{\mathrm{BS}},m}$ is defined in (5.26).

Proof. Please see Appendix C.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Remark 13. Since for $\mathrm{UE}_{n,1}$ the factor $\rho_{n,1}$ is one, we have $\bar{\mathbf{h}}_{n,1} = \tilde{\mathbf{h}}_{n,1}$. Thus, Theorem 4 is still valid for these users.

Remark 14. Theorem 5 states that the achievable rate of each user depends on the intra-cluster and inter-cluster misalignment factors, and a weak alignment reduces the power of the effective channel of that user. Intra-cluster and inter-cluster power allocation are other parameters that affect the achievable rate as seen in (5.25). Further, the bound shows that the maximum eigenvalue of the baseband precoder is important in maximizing the achievable rate. That is to say, the effective channel matrix should be designed in a way that the eigenvalues of the baseband precoder are as close as possible to each other. This is because if eigenvalues are far from each other, the maximum eigenvalue will be large. This increases the value of $\zeta_{\mathrm{inter}}^{n,m}$ which causes less achievable rate.

To gain some insight into the effect of beam misalignment, we extract a lower bound for the rate gap when $\mathrm{UE}_{n,m}$ receives the signal via LoS and NLoS channel.

Theorem 6. The rate gap between the perfect aligned and misaligned $\mathrm{UE}_{n,m}$ is given

by

$$\Delta R_{n,m} \triangleq \bar{R}_{n,m} - \tilde{R}_{n,m}$$

$$\leq \log_2 \left(1 + \frac{\left(1 - \rho_{n,m}^2\right) \kappa_{\max}(\mathbf{S}) + \sigma^2 K_{N_{\mathrm{BS}},m}^{-1} |\beta_{n,m}|^{-2}}{\rho_{n,m}^2 K_{N_{\mathrm{BS}},1}^{-1} \kappa_{\min}(\mathbf{F}) \sum_{k=1}^{m-1} P_{n,k}} \right). \qquad (5.27)$$

Proof. Please see Appendix C.4. ☐

The upper bound in Theorem 6 explicitly shows the effect of the parameters of HB-NOMA system on the rate performance. A low misalignment factor can substantially increase the rate gap.

Remark 15. In Section 5.3.1 the users are assumed to have LoS channels and to be perfectly aligned in a same direction. Particularly, Eq. (5.16) orders the users with respect to the their effective channel. Actually, these effective channels are the strongest path between the BS and users. However, when the users are not aligned in the same direction, the effective channels are not necessarily the strongest. This is because the users have to orient their antenna array response vector toward the beam direction of the first user rather than the best direction. Hence, to properly perform SIC, we revise the ordering considering the misalignment effective channel, i.e.,

$$\left\|\tilde{\mathbf{h}}_{n,1}\right\| \geq \left\|\tilde{\mathbf{h}}_{n,2}\right\| \geq \cdots \geq \left\|\tilde{\mathbf{h}}_{n,M_n}\right\|, \quad \text{for} \quad n = 1, 2, \ldots, N. \qquad (5.28)$$

Further, in (5.18) the aligned effective channel should be replaced by the misaligned effective channel.

5.5   Numerical Results

In this section we simulate the HB-NOMA system regarding the various design parameters to confirm the analytical derivations in Theorems 4-6. For simulations, since large scaling fading and path loss put more restriction on mmWave systems, the small scale fading is negligible. The defualt number of antennas N_{BS} N_{MU} for the BS and all users is assumed 32 and 8, respectively, unless it is mentioned. The misalignment is described as a random variable uniformly distributed by parameter b, i.e., $\varphi_{n,1} - \varphi_{n,m} \in [-b, b]$. We first present the results of the HB-NOMA with perfect alignment. Then, the effect of misalignment on the rate performance is shown. Finally, the sum-rate of HB-NOMA with OMA is illustrated.

5.5.1   Perfect Beam Alignment

Figure 5.3 studies the performance of the derived bound in Theorem 4 for aligned users. The users are not affected by the inter-cluster interference from other clusters. It is supposed that the number of users is two and channel gain of the strong and weak user is 0 and -2 dB, respectively. Fig. 5.3(a) reveals that the HB-NOMA approximately achieves the rate the same as that of fully-digital beamforming (FD beamforming) for a wide range of SNR. In particular, a small gap between the exact value of HB-NOMA and the lower bound is observed for the strong user ($\mathrm{UE}_{1,1}$). This is because the complicated expression of the noise term in (5.20) is replaced by a simple but greater term. For the weak user ($\mathrm{UE}_{1,2}$) the bound is very tight due to two reasons. First, in the SINR of the weak user, the noise term is dominated by the interference term. Therefore, the effect of noise term is neglected. Second, the interference term is modeled very accurately.
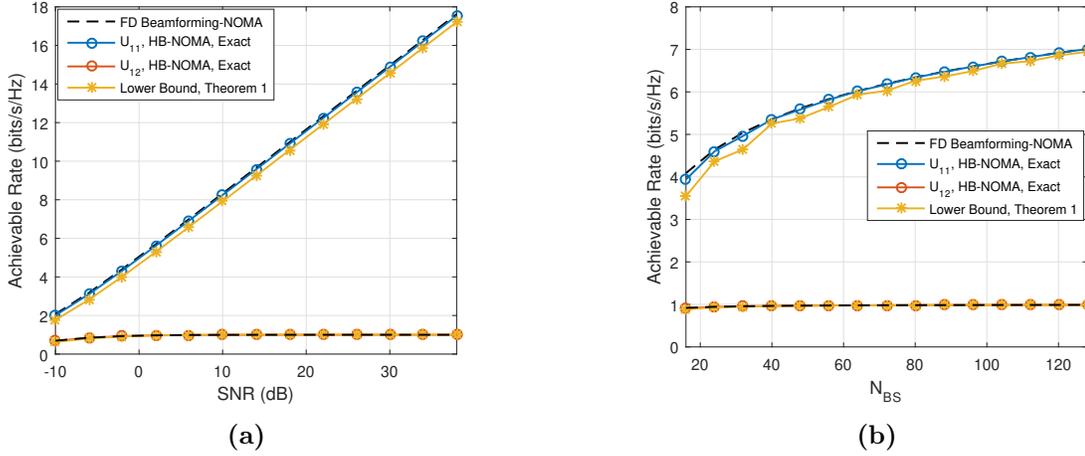
Figure 5.3   Evaluation of rate performance of the strong channel-based precoder in HB-NOMA with perfect alignment (LoS channels) in terms of (a) SNR and (b) N_{BS}.

Fig. 5.3(b) studies the achievable rate for various N_{BS}. For small N_{BS}s, the fully-digital outperforms the HB-NOMA. When N_{BS} is samll, the RF precoder is not able to steer a highly direct beam toward the users. By increasing N_{BS}, the beam becomes narrow and the users capture much more power. Again, for the weak user, the lower bound is accurate at all N_{BS} regions. For the strong user, the bound does not approach to the exact value but, for $N_{\text{BS}} > 60$, the bound is approximately the same as to the exact HB-NOMA.

5.5.2   Beam Misalignment

The beam misalignment effect is depicted by Figs. 5.4-5.7. We consider five clusters in which $\varphi_{1,1} = 10°$, $\varphi_{2,1} = 30°$, $\varphi_{3,1} = 50°$, $\varphi_{4,1} = 65°$, and $\varphi_{5,1} = 80°$. All simulations have been done for the middle cluster (third cluster) which likely imposed the same interference from all the other clusters. Also, the channel gain of the strongest user is 0 dB and the next user's gain drops 1 dB. For instance, the channel gain of $\text{UE}_{n,m}$
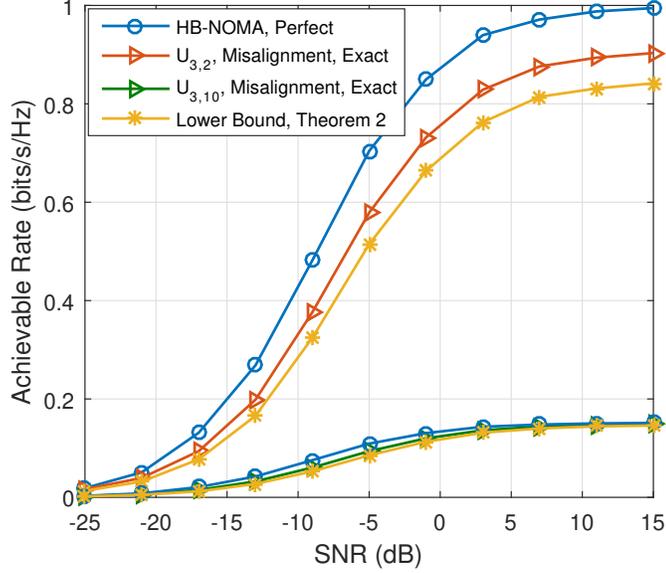
Figure 5.4   Evaluation of the misalignment on the rate performance of HB-NOMA versus SNR.

is $-(m-1)$ dB. In Figs. 5.4, 5.5, and 5.7, the number of users in the third cluster is 10.

In Fig. 5.4, the achievable rate of two misaligned users $UE_{3,2}$ (the strong user) and $UE_{3,10}$ (the weak user) versus SNR is shown where the channel gains are -1 and -9 dB, respectively. The misalignment parameter is assumed $b = 3$. The number of users in all the other clusters is equal to five. Two different observations are obtained. Increasing the SNR leads to a larger rate gap between perfectly aligned and the misaligned HB-NOMA for the strong user, whereas for the weak users both HB-NOMAs achieve almost the same rate for all SNRs. This demonstrates that the effect of misalignment on the strong users is greater than the weak users. In other words, the weak users should deal with the intra-cluster interference while the strong users should deal with the inter-cluster interference. The other observation is that the
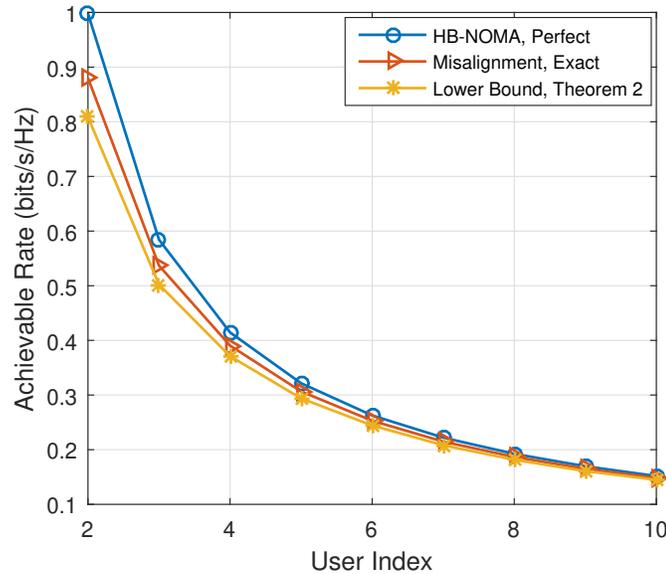
Figure 5.5   **Evaluation of the misalignment on the rate performance of HB-NOMA versus user index.**

lower bound is loose for the strong users but tight for the weak user. The observation indicates that our derived normalized effective channel model in Lemma 4 is precise for those users which are intra-cluster interference limited. That is, our finding is able to exactly model the intra-cluster interference. However, the loose lower bound for the strong user indicates that the inter-cluster interference is a little inaccurate which is due to approximating an $N-1$ dimensional subspace with one dimensional space provided in Appendix C.2.

To gain more details, we have simulated the achievable rate of all the misaligned users for SNR=15 dB in Fig. 5.5. Also, the number of users in the other clusters is set to 15. The mentioned two observations can be seen from this figure, t oo. However, for the strong user, the rate gap between the perfect HB-NOMA and misaligned HB-NOMA is smaller than that of Fig. 5.4. Another important observation gained
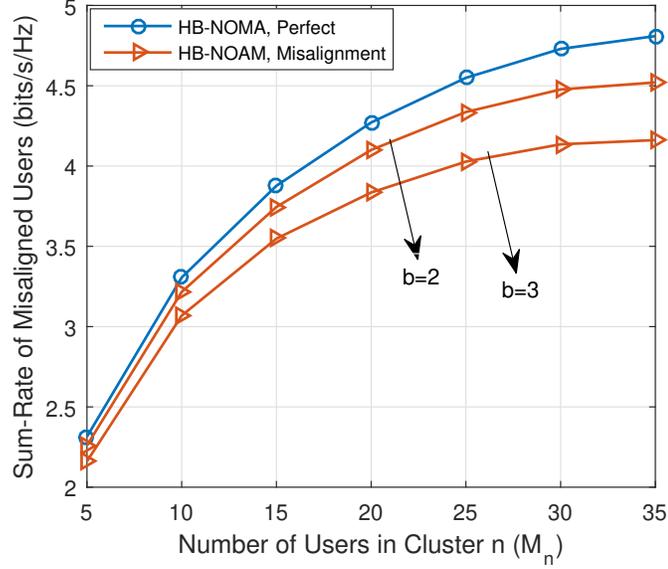
Figure 5.6: Evaluation of the misalignment on the rate performance of HB-NOMA versus number of users per cluster (M_n).

form Fig. 5.5 is the impact of the power allocation among the clusters. Based on the proposed power allocation scheme in (5.18), to achieve higher rate, more power is assigned to the other clusters than the third cluster which causes $UE_{3,2}$ to achieve the rate 0.91 bits/s/Hz. Whereas, for the previous scenario more power is allocated to the third cluster which has more users. Therefore, the rate of $UE_{3,2}$ is 0.88 bits/s/Hz. This shows that due to the misalignment the strong clusters leads to higher inter-cluster interference.

Fig. 5.6 compares the sum-rate performance of all the misaligned users with the perfectly aligned HB-NOMA users. Likewise Fig. 5.5, we set SNR=15 dB and 15 users for all the clusters except the third. The number of users in the third cluster varies from 5 to 35. Notice that the sum-rate is shown only for the misaligned users, e.g., the rate of the first user is neglected. By increasing the number of users, the allocated
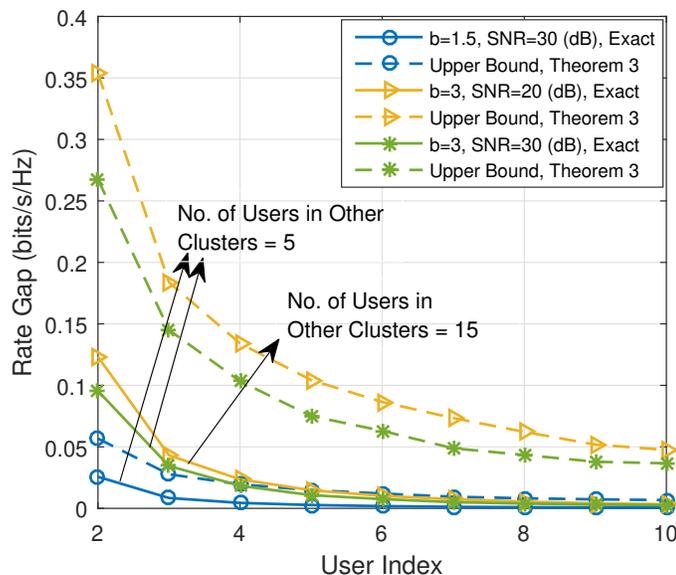
Figure 5.7   Demonstration of the rate gap among the different misaligned users.

power to the cluster increases. In consequence, the total rate increases. However, the difference between the aligned and misaligned HB-NOMA becomes worse. Although more users in a cluster means more power is allocated to, the number of users which have inter-cluster interference limited increases as well. As a result, it brings about a higher rate lost. Indeed, by making the misalignment parameter worse (b=6), the rate lost becomes bigger. It can be concluded that to avoid higher rate lost, HB-NOMA needs to schedule equal number of users per cluster to serve.

The upper bound evaluation for gap rate between the perfect alignment and misalignment is demonstrated by Fig. 5.7. The number of users in other clusters is 5 or 15. For SNR=30 dB and b=3, the gap is not substantial and the bound is close to the actual value. When b becomes larger, the gap between the stronger users is bigger than the weaker users. When number of the users of the other cluster increases and simultaneously SNR is reduced, only the stronger users' gap increases. To clarify,
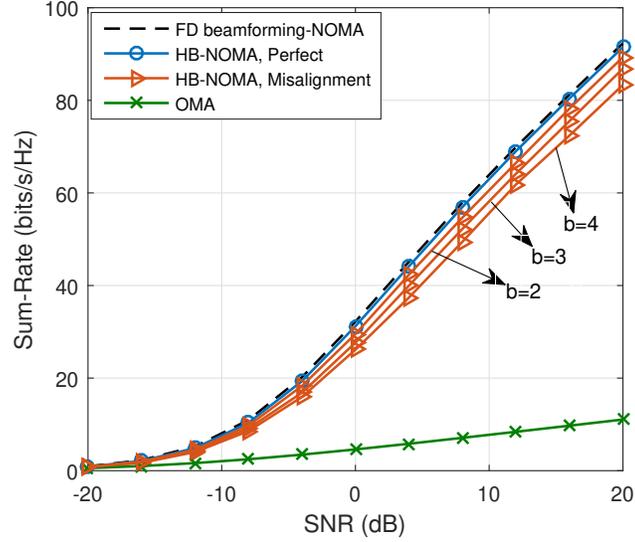
Figure 5.8: Sum-rate comparison of the three different systems. The fully-digital and hybrid beamforming systems serve the users using NOMA. The analog system supports the users by exploiting OMA.

for $UE_{3,2}$ to $UE_{3,5}$, the gap becomes larger, while for the remaining users it is unchanged. The bounds for $b=6$ are not very close to the exact rate gap curves. The main reason is that in the deriving process of the bound in the second line of (C.15) in Appendix C.4, the effect of the inter-cluster interference term is skipped. However, for high misalignment values the interference is considerable. This causes the extracted bound to be less accurate for higher misalignment.

Our HB-NOMA is compared with the traditional OMA technique in Fig. 5.8. We choose TDMA for OMA. To gain some insights, three different mmWave systems is evaluated. These systems are fully-digital beamforming, hybrid beamforming and analog beamforming. For fully-digital we assume $N_{BS} = N_{RF}=32$ which serve 8 clusters. Likewise, for hybrid beamforming we have $N_{BS}=32$ but $N_{RF}=8$. Both fully-digital and hybrid systems support 8 clusters of users. The first cluster has AoD of

$10°$ and AoD of the next clusters increases by $10°$. Further, the users inside of each clusters are distributed in a way that the maximum channel gain difference between the strongest and weakest user is 18 dB. Indeed, the channel gain of the strongest user is 0 dB. The first cluster contains 4 users and each next cluster serves two users more than the previous cluster. Totally, thanks to NOMA technique, both systems support 88 users in each time slot. For OMA, we assume the analog beamforming system equipped with only one RF chain is able to serve one user per time slot. For $UE_{n,m}$, the achievable rate of OMA is $\log_2(1 + P|\mathbf{w}_{n,m}\mathbf{H}_{n,m}\mathbf{f}_{RF}|^2/\sigma^2)$. As expected fully-digital NOMA system achieves the highest sum-rate performance. The HB-NOMA with perfect alignment achieves approximately the same rate as the full-digital. For $b{=}2$, the misaligned HB-NOMA shows a very close performance to the perfect HB-NOMA. By increasing b, the performance slightly decreases. There is a huge rate difference between HB-NOMA and OMA. We conclude that, even in the presence of misalignment, HB-NOMA outperforms OMA.

5.6   Summary and Conclusion

A hybrid beamforming-based NOMA has been designed for the downlink of a single-cell mmWave communication system. To study the achievable rate of an HB-NOMA user, we first formulated an optimization problem for the sum-rate of all users in the cell and then proposed an algorithm to solve it in three steps based on the strongest user precoder design. In order to evaluate the sum-rate, we found a lower bound for the achievable rate of each user under perfect and imperfect beam alignment between the effective channel of the users in each cluster. The lower bound analysis demonstrates that perfect HB-NOMA achieves a sum-rate close to that with fully-digital precoder. For the imperfect correlation, the relationship between the effective

channels of the first user and other users inside a cluster was modeled. The bound for the misalignment shows that it is highly function of the mislaigned angle. Such that, a large misalignment angle can cause a significant reduction in the achievable rate. Further, for each user, the rate gap between the perfect and imperfect alignment is bounded. The simulation results confirmed our findings.

CHAPTER 6:

MMWAVE HYBRID BEAMFORMING-NOMA SYSTEMS WITH BEAMWIDTH CONTROL

The use of directional transmission in millimeter-Wave (mmWave) frequencies results in limited channel coherence time. In this chapter, we take the limited channel coherence time into account for non-orthogonal multiple access (NOMA) in mmWave hybrid beamforming systems. Due to the limited coherence time, the beamwidth of the hybrid beamformer affects the beam-alignment time, which in turn directly impacts the data transmission rate. To investigate this trade-off, we utilize a combined beam-alignment algorithm. Then, we formulate a sum-rate expression which considers the channel coherence time and beam-alignment time as well as users' power and other system parameters. Further, a joint power and beamwidth optimization problem is solved by iterating between the power allocation and the beamwidth optimization. When allocating the power, we use the log-exponential reformulation and the sequential parametric convex approximation (SPCA) methods to solve the non-convex problem. Since beamwidth optimization involves too many variables, we propose an algorithm which iterates between clusters of users. Numerical results show that the optimized hybrid beamforming-NOMA (HB-NOMA) system can achieve much higher sum-rates compared to NOMA with analog beamforming and

traditional multiple access techniques.

6.1   Introduction

In practical scenarios, neglecting the effect of beam-alignment duration may cast doubt on the performance of the mmWave-NOMA networks. Especially, since the channel coherence time in mmWave bands is limited [21], the beam-alignment duration should be adequately small. Thus, on one hand, a small beam-alignment duration results in a wide beamwidth, i.e., low beamforming gain, and noisy channel estimation. On the other hand, a long beam-alignment time provides robust beamforming and accurate channel estimation but imposes a delay in data transmission. This may not be desirable in delay-sensitive systems as it leaves less time for data transmission and leads to low sum-rates. There is a rich literature on fast beam-alignment algorithms [22, 23, 79, 116–122]. This issue is very crucial in mmWave-NOMA networks in which more users are trained at each frequency/time resource. Beamwidth control and sum-rate trade-off in the mmWave analog beamforming-NOMA network for two users are evaluated in [123]. The impact of beam-alignment duration on the sum-rate of the system is determined and then an optimization problem that maximizes the sum-rate subject to the training duration and allocated power for each user is investigated. However, due to the inter-cluster interference, an extension of this architecture to the hybrid beamforming-NOMA (HB-NOMA) network is quite challenging.

6.1.1   Contributions

In this chapter, motivated by [123], we study the beamwidth control and sum-rate trade-off for the mmWave HB-NOMA network. There are two major differences between [123] and our work. First, we consider a hybrid beamforming system which

produces side lobes and as a result inter-cluster interference. Second, we do not allow receivers to have a beamwidth wider than that of their intended transmit beam. Otherwise, the receiver cannot catch the entire transmission energy. Neither the first case nor the second case is considered in [123]. The contributions of this chapter are listed below:

1. We consider the well-studied mmWave hybrid beamforming combined with NOMA for limited coherence time scenarios. The system can control the beamwidth, using the phase-shifters deployed in the hybrid beamformer, and allocate power to NOMA users. To this end, a tone-based beam-alignment algorithm [22] compatible to our mmWave-NOMA system is utilized. The algorithm combines the exhaustive search [23] and tone-based beam-alignment [22] algorithms.

2. Unlike the existing multi-beam mmWave-NOMA systems, we take the channel coherence time into account. The limited coherence time leads to a trade-off between the beamwidth resolution and the data transmission rate. We also formulate a new sum-rate expression for optimization.

3. A joint power and beamwidth optimization algorithm is proposed which iterates between the power allocation and beamwidth optimization.

4. The numerical results verify the effectiveness of the joint optimization algorithm. Also, two significant results are revealed. First, at low signal-to-noise ratios (SNRs), both power allocation and beamwidth control play a major role in the sum-rate while at high SNRs, beamwidth is the only important parameter. Second, a bottleneck for achieving high sum-rates is a small number of antennas, which results in a low resolution beamwidth, especially at large coherence time

and low SNRs.

The rest of the chapter is organized as follows. In Section 6.2, the system model is described. Section 6.3 formulates the optimization problem. In Section 6.4, the allocated power and beamwidth are determined through the proposed optimization algorithm and its convergence analysis is provided. Numerical results are presented in Section 6.5. Section 6.6 concludes the chapter.

Notations: Hereafter, $j = \sqrt{-1}$, small letters, bold letters, and bold capital letters designate scalars, vectors, and matrices, respectively. Superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ denote the transpose, conjugate, and transpose-conjugate operators, respectively. Further, $|\cdot|$, $\|\cdot\|$, $\|\cdot\|_2$, and $\|\cdot\|_F$ denote the absolute value, the norm-1, the norm-2, and the Frobenius norm, respectively. Also, $\mathbb{E}(\cdot)$ denotes the expectation. Finally, $\lceil \cdot \rceil$ denotes the ceiling function.

6.2   System Model

We consider a similar system model described in Section 5.2. Also, we adopt the channel model in (2.1). In what follows, we explain more details about the channel model. In mmWave outdoor communications, to further reduce the interference, sectorized BSs can be employed. Mostly, each sector in the azimuth domain is much wider than that of the elevation domain [32, 34]. Reasonably, we assume that the transmitter separates the clusters in the azimuth domain and considers fixed elevation angles. Further, we assume that the sector-level beamwidth for the transmitter is defined by ω_t and for each user is defined by ω_r. Hence, the transmitter implements only azimuth beamforming and neglects elevation beamforming. In this case, the antenna configuration is a uniform linear array (ULA) and the superscript "el" is

dropped. For a ULA, the steering vector is defined as

$$\mathbf{a}_\text{t}(\phi_l) = \left[1, e^{-j\pi\phi_l}, \ldots, e^{-j\pi(N_\text{t}-1)\phi_l}\right]^T, \tag{6.1}$$

where $\phi_l \in [-1, 1]$ is related to the AoD $\varphi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ as $\phi_l = \frac{2D\sin(\varphi)}{\lambda}$. Note that D denotes the antenna spacing and λ denotes the wavelength of the propagation. The antenna array response vector for $\mathbf{a}_\text{r}(\theta_l)$ can be written in a similar fashion.

AoD/AoA variations over the coherence time are trivial and can be ignored [124]. Let T and T_b denote the coherence time and the time duration over which AoD/AoA remain unchanged, respectively. In [124], it is shown that the coherence time duration is far less than T_b, i.e., $T \ll T_b$, which ensures that AoD/AoA do not change over the coherence time. Here, the channel gain captures path loss and shadow fading effects. The assumption on AoD/AoA and the channel gain state that the channel model in (2.1) represents a long term channel which is widely adopted in the literature [116, 122, 124]. Ignoring AoD/AoA variations and instantaneous channel fluctuations are valid assumptions since the power allocation and the beamwidth control are done over the coherence time. This follows from the fact that the long term channel model can be effectively used in long term resource allocation [124].

It is demonstrated that in dense urban environments, with high probability, the mmWave channels contain only one or two paths, with the dominant one that carries most of the signal energy [67]. Therefore, with a single path assumption, the MPC model described in (2.1) is converted to a single path channel model given by

$$\mathbf{H}_{n,m} = \beta_{n,m}\mathbf{a}_\text{UE}(\theta_{n,m})\mathbf{a}_\text{BS}^\dagger(\phi_{n,m}). \tag{6.2}$$

Hence, the BS communicates to the users through a single path channel. It is worth mentioning that the users are ordered based on their channel gain, i.e., $\beta_{n,1} \geq, \ldots, \geq \beta_{n,M}$ where $\beta_{n,m}$ is captured through channel quality indicator (CQI) [125]. Although it is assumed that the channel is single path, in some rare cases there might be more than one dominant path. To mitigate the multipath issue, rake receivers or orthogonal frequency-division multiplexing (OFDM) can be used. It should also be mentioned that due to the availability of large bandwidth, in mmWave systems, wide band transmission is preferred. For this case, the considered narrow band system should be combined with OFDM. In general, the extension of our narrow band system to the wide band is straightforward and studied in the literature. For instance, the OFDM-based NOMA has been considered in [54] and other similar work.

6.2.1   Beam-alignment

Each transmission frame in mmWave directional communications depends on the channel coherence time and consists of two parts: (i) beam-alignment and (ii) data transmission as depicted in Fig. 6.1. At the first step, the channel parameters AoDs, AoAs, and effective channel are estimated by channel estimation algorithms. In this chapter, we assume that the channel parameters are perfectly estimated [116, 123]. In particular, the estimation of AoDs and AoAs is performed using beam-alignment algorithms and takes much more time compared to the effective channel estimation. The beam-alignment algorithms should be fast, accurate, and energy-efficient. At the second step, during the remaining time, the data is transmitted. Recently, a few codebook-based beam-alignment algorithms have been proposed for mmWave hybrid beamforming systems [22, 23, 79, 117, 125]. Even the current fastest algorithms take a considerable portion of the coherence time that leaves a short time for data
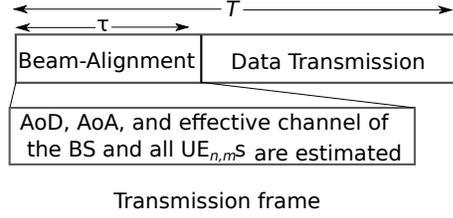
Figure 6.1   Schematic of the transmission frame in the mmWave-NOMA system.

transmission and can diminish the achievable rate of a user [126]. On the other hand, a smaller beam-alignment duration means wider beamwidth, which supplies lower beamforming gain. Consequently, in the mmWave systems, there exists a trade-off between the training duration and data transmission duration. This trade-off becomes more notable in the mmWave-NOMA networks in which more channels should be estimated. Motivated by this, finding an optimal beam-alignment duration and user power allocation for data transmission to increase the sum-rate of the mmWave-NOMA system will be subject of this chapter.

As mentioned before, the main part of a beam-alignment algorithm consists of beam-alignment followed by an effective channel estimation. In general, there are two different search algorithms for beam-alignment, exhaustive search [23] and hierarchical search [79]. The former algorithm examines all beam pairs in the codebook for BS and UE and determines the best pair that maximizes the beamforming gain. The training time for this algorithm is proportional to the size of the beam's search space which is given by

$$\tau = \left\lceil \frac{\omega_{\mathrm{BS}}}{\eta} \right\rceil \left\lceil \frac{\omega_{\mathrm{UE}}}{\mu} \right\rceil T_p, \tag{6.3}$$

where η and μ denote the beamwidths of the BS and the UE, respectively. Further, T_p is the time for pilot transmission. On the other hand, the hierarchical search algorithm is designed based on multi-level codebook designs and uses bisection beam

search. At the first level, the algorithm chooses a wider beam with a low resolution which has a small beam search space. The algorithm refines the search iteratively using the next-level codebook within the subspace defined in the wider-level. At each level, the algorithm performs an exhaustive search to find the best pair. Compared to the exhaustive search, the hierarchical algorithm takes less training time with the same beam resolution and length of the pilot sequence at the cost of the higher probability of misalignment [125].

The exhaustive and hierarchical algorithms are designed only for single-user and multi-user scenarios. In particular, multi-user beam-alignment algorithms assume that each user has a distinct AoD and might not be efficient for NOMA systems in which users are allowed to have the same AoD. Particularly, the hierarchical algorithm has a higher probability of beam misalignment at the low SNR regime [125]. This can be a major barrier in realizing the hierarchical algorithm in mmWave-NOMA networks where the users with low SNR are paired with the users with high SNR. It seems that the exhaustive algorithm is a proper candidate for the beam-alignment in the mmWave-NOMA system since it works better at the low SNR regime [125]. In the exhaustive search algorithm, all beams are aligned with the same resolution. That is, the beamwidth of the beams at the BS is equal. In some scenarios, this may impose a limitation on designing an optimal mmWave-NOMA system. To overcome this issue, we adopt a multi-user tone-based beam-alignment algorithm proposed in [22]. The algorithm consists of three steps summarized as follows. At the first step, each user transmits a pilot using one omni-directional antenna with a unique frequency tone in the uplink. Given a predefined resolution $\frac{\omega_{\text{BS}}}{\eta}$ for each user, the BS searches for the best AoD that maximizes the beamforming gain. It is worth mentioning that

the BS can estimate the AoDs with different predefined resolutions. At the second step, using the estimated AoDs, the BS simultaneously transmits a pilot for each user over a unique frequency tone in the downlink. Each user estimates the AoA with a predefined resolution $\frac{\omega_{UE}}{\mu}$. Finally, each user transmits an orthogonal pilot sequence to the BS, and the BS estimates the channel. We note that using a unique tone for each user requires more hardware complexity compared to the search algorithm. There are two main differences between the tone-based algorithm and the exhaustive search algorithm. First, due to using unique frequency tones, the beam-alignment for each user is done independently. Hence, the BS can select different beamwidth values for different users, and each user can also have a distinct beamwidth value. Second, the beam-alignment time is shorter than those of the exhaustive search algorithm. That is, the total training time for the tone-based algorithm is $\tau = \max\{(\lceil \frac{\omega_{BS}}{\eta_n} \rceil + \lceil \frac{\omega_{UE}}{\mu_{n,m}} \rceil) T_p\}$, where η_n and $\mu_{n,m}$ denote the beamwidth of the nth beam of the BS and the $UE_{n,m}$, respectively. It is clear when η_n and $\mu_{n,m}$ are the same as those of the exhaustive search algorithm, the training time for the tone-based algorithm is smaller than (6.3). It is worth mentioning that user clustering is another important topic that lies in the beam-alignment phase but out of the scope of this work.

Although the algorithm in [22] is applicable to the mmWave-NOMA structure and can remarkably reduce the training time, similar to the hierarchical algorithm, it may result in a higher probability of misalignment. This is due to the use of omni-directional antennas at the first step which does not provide enough beamforming gain, especially for low-SNR users. To tackle this challenge, we modify the algorithm at the cost of sacrificing the speed of beam-alignment. We assume that each user steers directional beams with the predefined beamwidth μ. Then, we combine the first

and second steps and perform an exhaustive beam search to find the best beam pair that achieves the highest beamforming gain. Note that the BS communicates with each user via a unique frequency tone. Further, the third step remains unchanged. Therefore, the training time becomes $\tau = \max\{\lceil\frac{\omega_{\text{BS}}}{\eta_n}\rceil\lceil\frac{\omega_{\text{UE}}}{\mu_{n,m}}\rceil T_p\}$. When beamwidth for the BS and users are the same as those of (6.3), the beam-alignment time for the modified tone-based algorithm is similar to that of the exhaustive search algorithm. In summary, we adopt the tone-based beam-alignment algorithm in [22] and instead of hierarchical search we use exhaustive search. It is worth mentioning that user clustering is another important topic that lies in the beam-alignment phase but is out of the scope of this work.

6.2.2   Hybrid Beamforming Gain and SINR

Let us define $\mathbf{f}_n = \mathbf{F}_{\text{RF}}\mathbf{f}_{\text{BB}}^n$ as the hybrid beamforming vector of the nth beam at the BS. An ideal hybrid beamformer leads to $|\mathbf{a}_{\text{BS}}^{\dagger}(\phi_{n,m})\mathbf{f}_n| = \sqrt{G_{\text{BS}}^{\text{id}}(\phi'_{n,m},\eta_n)}$ in which $G_{\text{BS}}^{\text{id}}$ is the beamforming gain of the ideal beamformer at the BS and $\phi'_{n,m}$ and η_n denote the angle off the boresight direction at the BS and the beamwidth of the nth beam, respectively. It is worth mentioning that in this chapter the parameter $G_{\text{BS}}^{\text{id}}$ is irrespective of how the hybrid beamforming is designed. Essentially, the value of $G_{\text{BS}}^{\text{id}}$ depends on the beamformer $\mathbf{f}_n$, where $||\mathbf{f}_n|| = 1$, and the size of the transmit antenna array. Also, note that an ideal beamforming vector is obtained when there is no channel estimation error and perfect beam-alignment is done while considering an infinite resolution for the phase-shifters. Further, the beamwidth depends on the design of the analog beamformer and the digital beamformer. In particular, the
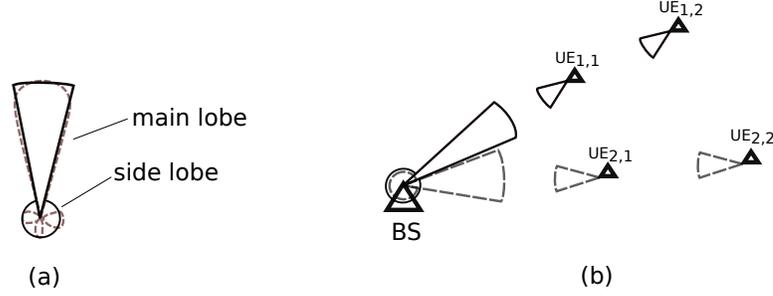
Figure 6.2 **(a) A non-ideal beam is modeled with a constant main lobe gain and side lobe gain, (b) The impact of the side lobe gain of each beam on the UEs located in other clusters.**

beamforming gain is defined as

$$
G_{\mathrm{BS}}^{\mathrm{id}}(\phi_{n,m}', \eta_n) = \begin{cases} \frac{2\pi}{\eta_n}, & \text{if } |\phi_{n,m}| \leq \frac{\eta_n}{2}, \\ 0, & \text{otherwise.} \end{cases}
\tag{6.4}
$$

Further, the beamforming gain of the ideal analog beamformer at $\mathrm{UE}_{n,m}$ is assumed to be $|\mathbf{w}_{n,m}^{\dagger}\mathbf{a}_{\mathrm{UE}}(\theta_{n,m})| = \sqrt{G_{\mathrm{UE}}^{\mathrm{id}}(\theta_{n,m}', \mu_{n,m})}$ in which $G_{\mathrm{UE}}^{\mathrm{id}}$ is the gain of the ideal analog beamformer and $\theta_{n,m}'$ and $\mu_{n,m}$ respectively denote the angle off the boresight direction at $\mathrm{UE}_{n,m}$ and the beamwidth of $\mathrm{UE}_{n,m}$. Similar to $G_{\mathrm{BS}}^{\mathrm{id}}$, the ideal beamforming gain is defined as

$$
G_{\mathrm{UE}}^{\mathrm{id}}(\theta_{n,m}', \mu_{n,m}) = \begin{cases} \frac{2\pi}{\mu_{n,m}}, & \text{if } |\theta_{n,m}'| \leq \frac{\mu_{n,m}}{2}, \\ 0, & \text{otherwise.} \end{cases}
\tag{6.5}
$$

Note that ideal beamforming at the BS and users results in the complete cancellation of the inter-cluster interference represented in (5.6).

In practice, achieving the ideal beamforming gain may not be possible because of the application of suboptimal solutions for the analog beamformer, finite resolution of the phase-shifters, channel estimation error, and beam misalignment. These problems

reduce the gain in the main lobe and introduce a side lobe for each beam. Hence, the beamforming model should take these issues into account [79, 105]. A more practical model for the beamforming gain of the nth beam is given by

$$G_{\text{BS}}(\phi'_{n,m}, \eta_n) = \begin{cases} \frac{2\pi - (2\pi - \eta_n)\xi}{\eta_n}, & \text{if } |\phi'_{n,m}| \leq \frac{\eta_n}{2}, \\ \xi, & \text{otherwise}, \end{cases} \tag{6.6}$$

where $0 \leq \xi < 1$ with $\xi \ll 1$ for narrow beams, which is widely adopted in the literature [106, 117, 120]. Note that there is another common model for the beamforming gain with side lobe level varying with the beamwidth [105]. In this chapter, to make the analysis tractable, we use the model described above that satisfies the total power of 1, i.e., $\int_0^{2\pi} G_{\text{BS}}(\phi'_{n,m}, \eta_n)\mathrm{d}\phi'_{n,m} = \frac{\eta_n}{2\pi}\frac{2\pi - (2\pi - \eta_n)\xi}{\eta_n} + \frac{2\pi - \eta_n}{2\pi}\xi = 1$. Similarly, the model for the beamforming gain of $\text{UE}_{n,m}$ is given by

$$G_{\text{UE}}(\theta'_{n,m}, \mu_{n,m}) = \begin{cases} \frac{2\pi - (2\pi - \mu_{n,m})\xi}{\mu_{n,m}}, & \text{if } |\theta'_{n,m}| \leq \frac{\mu_{n,m}}{2}, \\ \xi, & \text{otherwise}. \end{cases} \tag{6.7}$$

In the above equations, the main lobe's gain is distributed uniformly in the entire beamwidth and the side lobe's gain is assumed to be constant [22, 118] as demonstrated in Fig. 6.2.(a). However, in reality, the main lobe's gain changes over the beamwidth and the side lobe's gain depends on the size of the beamwidth. For example, for a narrower beam, the side lobe's gain is higher [127]. Further, the side lobe results in interference that impacts the UEs located in other clusters as shown in Fig. 6.2.(b). In our formulation, this interference is modeled by the inter-cluster interference term in (5.6).

Hence, using (6.6), and (6.7), the signal-to-interference-plus-noise ratio (SINR) of $\text{UE}_{n,1}$ in the nth beam is expressed as

$$\gamma_{n,1} = \frac{P_{n,1}\beta_{n,1}^2 G_{\text{BS}}(\phi'_{n,1},\eta_n)G_{\text{UE}}(\theta'_{n,1},\mu_{n,1})}{\sum\limits_{q\neq n}^{N_{\text{RF}}}\sum\limits_{\ell=1}^{K_q} P_{q,\ell}\beta_{n,1}^2 G_{\text{UE}}(\theta'_{n,1},\mu_{n,1})\xi + \sigma^2}, \tag{6.8}$$

and, using (6.6) and (6.7), the SINR of $\text{UE}_{n,m}$, $m > 1$, is given by

$$\gamma_{n,m} = \frac{P_{n,m}\beta_{n,m}^2 G_{\text{BS}}(\phi'_{n,m},\eta_n)G_{\text{UE}}(\theta'_{n,m},\mu_{n,m})}{\sum\limits_{k=1}^{m-1} P_{n,k}\beta_{n,m}^2 G_{\text{BS}}(\phi'_{n,m},\eta_n)G_{\text{UE}}(\theta'_{n,m},\mu_{n,m}) + \sum\limits_{q\neq n}^{N_{\text{RF}}}\sum\limits_{\ell=1}^{K_q} P_{q,\ell}\beta_{n,m}^2 G_{\text{UE}}(\phi'_{n,m},\mu_{n,m})\xi + \sigma^2}. \tag{6.9}$$

Due to the single cell assumption, we can conclude that the users do not receive any interference from the side lobe and only receive signal from the main lobe. Further, it is assumed that codebooks for a specific level (beam resolution) are designed efficiently such that the steered beams by the BS do not overlap [23]. Hence, each user receives the desired signal and intra-cluster interference sent through the main lobe of the desired beam and inter-cluster interference sent through the side lobe of the other beams as visualized in Fig. 6.2.(b). Further, in the modified beam-alignment algorithm described in Section 6.2.1 and adopted in this section, the beams directed by the BS can have different beamwidth values, i.e., the beams have different resolutions. In this case, during the data transmission, beam overlap may occur, which can impose severe inter-cluster interference. To avoid this, we assume that there is a proper angle gap between the two neighboring beams. In mmWave hybrid beamforming, a limited number of RF chains is used, i.e., the number of beams is limited [20]. Therefore, the direction of each beam is selected to satisfy the required angle gap between the

beams. Further, each cluster's users are served via a common beam directed by the BS. Therefore, the training time for $\text{UE}_{n,m}$ is given by $\tau_{n,m} = \lceil \frac{\omega_{\text{BS}}}{\eta_n} \rceil \lceil \frac{\omega_{\text{UE}}}{\mu_{n,m}} \rceil T_p$ as explained in Section 6.2.1. Accordingly, the achievable rate for $\text{UE}_{n,m}$ can be calculated as

$$R_{n,m} = \left(1 - \frac{\tau}{T}\right) \log_2(1 + \gamma_{n,m}), \tag{6.10}$$

where T denotes the channel coherence time as indicated in Fig. 6.1. It is worth mentioning that the chosen frame duration is smaller than the channel coherence time.

6.3 Problem Formulation

Here, NOMA is performed for two UEs per cluster which is compatible with the multi-user superposition transmission schemes recently adopted by 3GPP [57, 128]. Further, the BS is assumed to generate only two beams. Extension to more than two clusters will be addressed in future work. To optimize the sum-rate performance, η_n, $\mu_{n,m}$, and $P_{n,m}$ should be optimized according to

$$\underset{\boldsymbol{\eta}, \boldsymbol{\mu}, \mathbf{P}}{\text{maximize}} \quad \sum_{n=1}^{2} \sum_{m=1}^{2} R_{n,m} \tag{6.11a}$$

$$\text{subject to} \quad \eta_{\min} \leq \eta_n \leq \omega_{\text{BS}}, \tag{6.11b}$$

$$\mu_{\min} \leq \mu_{n,m} \leq \min\{\omega_{\text{UE}}, \eta_n\}, \tag{6.11c}$$

$$\tau \leq T, \tag{6.11d}$$

$$\tau \geq \frac{\omega_{\text{BS}}}{\eta_n} \frac{\omega_{\text{UE}}}{\mu_{n,m}} T_p, \tag{6.11e}$$

$$R_{n,m} \geq R_{\min}, \tag{6.11f}$$

$$\sum_{n=1}^{2} \sum_{m=1}^{2} P_{n,m} \leq P_{\text{tot}}, \tag{6.11g}$$

$$P_{n,m} > 0, \tag{6.11h}$$

where $\boldsymbol{\eta} = [\eta_1, \eta_2]$, $\boldsymbol{\mu} = [\mu_{1,1}, \mu_{1,2}, \mu_{2,1}, \mu_{2,2}]$, $\mathbf{P} = [P_{1,1}, P_{1,2}, P_{2,1}, P_{2,2}]$, and P_{tot} denotes the total power of the BS. The smallest beamwidth resolutions for the BS and UE are denoted by $\eta_{\min}$ and $\mu_{\min}$, respectively. Here, we assume $\eta_{\min} = \mu_{\min}$. The beamwidth resolution relates to the number of antennas. Usually, the number of antennas at a BS is larger than those of UEs. Thus, the BS can generate narrower beams. However, we assume that the minimum beamwidths of the BS and UEs are identical. For the sake of simplicity, we relax $\tau \geq \lceil \frac{\omega_{\text{BS}}}{\eta_n} \rceil \lceil \frac{\omega_{\text{UE}}}{\mu_{n,m}} \rceil T_p$ to $\tau \geq \frac{\omega_{\text{BS}}}{\eta_n} \frac{\omega_{\text{UE}}}{\mu_{n,m}} T_p$ in (6.11e). After we obtain the optimal $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$, we can recalculate $\tau = \max\{\lceil \frac{\omega_{\text{BS}}}{\eta_n} \rceil \lceil \frac{\omega_{\text{UE}}}{\mu_{n,m}} \rceil T_p\}$.

6.4 Joint beamwidth control and power allocation

Problem (6.11) is an intractable non-convex optimization problem and needs to be decomposed. We propose an algorithm which iterates between the power allocation

and the beamwidth optimization. When allocating the power, we fix the beamwidths and when optimizing the beamwidths, we keep the powers fixed. We assume that the BS and the users are aligned after the training process, which means $|\phi_{n,m}| \leq \frac{\eta_n}{2}$ and $|\theta_{n,m}| \leq \frac{\mu_n}{2}$. We also assume the users within the same cluster have the same beamwidth, i.e., $\mu_{n,1} = \mu_{n,2} = \mu_n$. However, the users in different clusters do not necessarily have the same beamwidth.

6.4.1 Power allocation

When the beamwidth and the training time are fixed, the beamforming gains are also fixed. Then, problem (6.11) is simplified to:

$$\underset{\mathbf{P}}{\text{maximize}} \quad \sum_{n=1}^{2}\sum_{m=1}^{2} R_{n,m} \tag{6.12a}$$

$$\text{subject to} \quad (6.11f)-(6.11h). \tag{6.12b}$$

Although Problem (6.12) has been greatly simplified compared to Problem (6.11), its objective is still complicated and non-convex. To transform Problem (6.12) into a tractable form, we use the log-exponential reformation idea in [129]. Introducing slack variables $\boldsymbol{S} = [x_{n,m}, d_{n,m}], n = 1,2, m = 1,2$, we can transform the objective in Problem (6.12) into a linear form by $\sum_{n=1}^{2}\sum_{m=1}^{2} \log_2 \frac{2^{x_{n,m}}}{2^{d_{n,m}}}$. For the sake of brevity, we denote $G_{\text{BS}}(\eta_n)G_{\text{UE}}(\mu_n)$ by G_n and $G_{\text{UE}}(\mu_n)$ by G_{UE}^n and make the following definitions:

$$\text{SI}_{n,1} \triangleq P_{n,1}\beta_{n,1}^2 G_n + \sum_{\ell=1,q\neq n}^{2} P_{q,\ell}\beta_{n,1}^2 G_{\text{UE}}^n \xi + \sigma^2, \tag{6.13}$$

$$\text{SI}_{n,2} \triangleq P_{n,2}\beta_{n,2}^2 G_n + P_{n,1}\beta_{n,2}^2 G_n + \sum_{\ell=1,q\neq n}^{2} P_{q,\ell}\beta_{n,2}^2 G_{\text{UE}}^n \xi + \sigma^2, \tag{6.14}$$

$$\mathrm{I}_{n,1} \triangleq \sum_{\ell=1, q\neq n}^{2} P_{q,\ell} \beta_{n,1}^2 G_{\mathrm{UE}}^n \xi + \sigma^2, \tag{6.15}$$

$$\mathrm{I}_{n,2} \triangleq P_{n,1} \beta_{n,2}^2 G_n + \sum_{\ell=1, q\neq n}^{2} P_{q,\ell} \beta_{n,2}^2 G_{\mathrm{UE}}^n \xi + \sigma^2. \tag{6.16}$$

Then, Problem (6.12) can be rewritten as

$$\underset{\boldsymbol{S}, \mathbf{P}}{\mathrm{maximize}} \qquad (1 - \frac{\tau}{T}) \sum_{n=1}^{2} \sum_{m=1}^{2} (x_{n,m} - d_{n,m}) \tag{6.17a}$$

$$\mathrm{subject\ to} \qquad 2^{x_{n,1}} \leq \mathrm{SI}_{n,1}, \tag{6.17b}$$

$$2^{x_{n,2}} \leq \mathrm{SI}_{n,2}, \tag{6.17c}$$

$$2^{d_{n,1}} \geq \mathrm{I}_{n,1}, \tag{6.17d}$$

$$2^{d_{n,2}} \geq \mathrm{I}_{n,2}, \tag{6.17e}$$

$$(1 - \frac{\tau}{T})(x_{n,m} - d_{n,m}) \geq R_{\min}, \tag{6.17f}$$

$$(6.11g) - (6.11h). \tag{6.17g}$$

In Problem (6.17), the optimum is achieved when the constraints (6.17b)-(6.17e) satisfy with equality. Let us use (6.17b) as an example to show that the equality should be satisfied at the optimum. Assuming the opposite, we can increase $x_{n,1}$ while keeping other variables fixed. This results in increasing the cost function and contradicts the optimality assumption. Since constraints (6.17b)-(6.17e) achieve equality at the optimum, the non-convex objective of Problem (6.12) is equivalently decomposed into (6.17a) and constraints (6.17b)-(6.17e).

Unfortunately, constraints (6.17d) and (6.17e) are still non-convex. To relax the non-convex constraint to convex constraints, we use a sequential parametric convex approximation method (SPCA) [130]. In this method, the non-convex feasible set is sequentially approximated by an inner convex approximation. Using (6.17d) as an example, at Iteration k, since function $2^{d_{n,1}}$ is a convex function, i.e., $2^y - 2^x \geq 2^x \log 2(y - x)$, we have a lower bound of $2^{d_{n,1}}$ as:

$$2^{d_{n,1}^*[k-1]} \log 2(d_{n,1} - d_{n,1}^*[k-1]) + 2^{d_{n,1}^*[k-1]} \leq 2^{d_{n,1}}, \tag{6.18}$$

where $d_{n,1}^*[k-1]$ is the optimal solution at Iteration $k - 1$. Based on (6.18), we can relax (6.17d) into a convex constraint as

$$2^{d_{n,1}^*[k-1]} \log 2(d_{n,1} - d_{n,1}^*[k-1]) + 2^{d_{n,1}^*[k-1]} \geq \sum_{\ell=1,q \neq n}^{2} P_{q,\ell} \beta_{n,1}^2 G_{\text{UE}}^n \xi + \sigma^2. \tag{6.19}$$

Using the same method for (6.17e), we have

$$2^{d_{n,2}^*[k-1]} \log 2(d_{n,2} - d_{n,2}^*[k-1]) + 2^{d_{n,2}^*[k-1]} \geq P_{n,1} \beta_{n,2}^2 G_n + \sum_{\ell=1,q \neq n}^{2} P_{q,\ell} \beta_{n,2}^2 G_{\text{UE}}^n \xi + \sigma^2. \tag{6.20}$$

At each iteration, we can relax Problem (6.17) into the following convex problem:

$$\underset{\boldsymbol{S}, \mathbf{P}}{\text{maximize}} \qquad (1 - \frac{\tau}{T}) \sum_{n=1}^{2} \sum_{m=1}^{2} (x_{n,m} - d_{n,m}) \qquad (6.21a)$$

$$\text{subject to} \qquad (6.17b), \qquad (6.21b)$$

$$(6.17c), \qquad (6.21c)$$

$$(6.19) - (6.20), \qquad (6.21d)$$

$$(6.11g) - (6.11h). \qquad (6.21e)$$

This is a convex problem, which can be efficiently solved by off-the-shelf solutions, such as CVX [131].

By relaxing Problem (6.17) to Problem (6.21) in each iteration, we can propose an iterative algorithm to provide an approximation solution for Problem (6.17). Detailed steps are presented in Alg. 1. According to [130], Alg. 1 converges.

Algorithm 1 Power allocation

1: Set the sum-rate $R_{\text{sum}}[-1] \quad 0$, the maximal iteration number $k_{max} \quad 1000$ and the convergence threshold $\epsilon \quad 10^{-3}$;
2: **repeat**
3:     Choose a feasible start point $\mathbf{P}^*[0]$;
4:     $x_{n,m}^*[0] \quad \log_2(P_{n,m}^*[0]\beta_{n,m}^2 G_n)$,
5:     $d_{n,1}^*[0] \quad \log_2(\sum_{\ell=1, q \neq n}^{2} P_{q,\ell}^*[0]\beta_{n,1}^2 G_{\text{UE}}^n \xi + \sigma^2)$;
6:     $d_{n,2}^*[0] \leftarrow \log_2(P_{n,1}^*[0]\beta_{n,2}^2 G_n + \sum_{\ell=1, q \neq n}^{2} P_{q,\ell}^*[0]\beta_{n,2}^2 G_{\text{UE}}^n \xi + \sigma^2)$;
7: **until** (6.17f) is satisfied
8: $k \quad 0$;
9: **while** $R_{\text{sum}}[k] - R_{\text{sum}}[k-1] \geq \epsilon R_{\text{sum}}[k-1]$ and $k \leq k_{max}$ **do**
10:     $k \leftarrow k + 1$;
11:     Solve (6.21) to obtain $\boldsymbol{S}^*[k]$ and $\mathbf{P}^*[k]$;
12: **end while**
13: Return $\mathbf{P}^*[k]$.

6.4.2 Beamwidth optimization

When the powers are fixed, the problem to optimize the beamwidth can be rewritten
as

$$\underset{\boldsymbol{\eta}, \boldsymbol{\mu}}{\text{maximize}} \quad \sum_{n=1}^{2} \sum_{m=1}^{2} R_{n,m} \tag{6.22a}$$

$$\text{subject to} \quad (6.11b) - (6.11f). \tag{6.22b}$$

Problem (6.22) is a very complicated problem with non-convex objective and con-
straints. To simplify the problem, we first perform the following variable substitu-
tions:

$$\mu_n = \frac{A}{f_n}, \tag{6.23}$$

$$\eta_n = \frac{A}{h_n}, \tag{6.24}$$

where $f_n \triangleq G_{\text{UE}}(\mu_n) - \xi$, $h_n \triangleq G_{\text{BS}}(\eta_n) - \xi$ and $A = 2\pi - 2\pi\xi$ is a constant. By assum-
ing $G_{\text{BS}}(\eta_n) G_{\text{UE}}(\mu_n) \approx \frac{(2\pi - 2\pi\xi)^2}{\eta_n \mu_n} + \xi^2$, we have $\eta_n \mu_n = \frac{A^2}{h_n f_n}$ and $G_{\text{BS}}(\eta_n) G_{\text{UE}}(\mu_n) = h_n f_n + \xi^2$. Then, the SINR for $\text{UE}_{n,1}$ and $\text{UE}_{n,2}$ can be rewritten as

$$\gamma_{n,1} = \frac{P_{n,1} \beta_{n,1}^2 (h_n f_n + \xi^2)}{\sum_{\ell=1, q \neq n}^{2} P_{q,\ell} \beta_{n,1}^2 (f_n + \xi)\xi + \sigma^2}, \tag{6.25}$$

$$\gamma_{n,2} = \frac{P_{n,2} \beta_{n,2}^2 (h_n f_n + \xi^2)}{P_{n,1} \beta_{n,2}^2 (h_n f_n + \xi^2) + \sum_{\ell=1, q \neq n}^{2} P_{q,\ell} \beta_{n,2}^2 (f_n + \xi)\xi + \sigma^2}. \tag{6.26}$$

Instead of finding the optimal beamwidth, we find the optimal f_n and h_n. We can

rewrite Problem (6.22) as

$$\underset{\boldsymbol{h}, \boldsymbol{f}}{\text{maximize}} \quad \sum_{n=1}^{2} \sum_{m=1}^{2} R_{n,m} \tag{6.27a}$$

$$\text{subject to} \quad \frac{A}{\omega_{\text{BS}}} \leq h_n \leq \frac{A}{\eta_{\min}}, \tag{6.27b}$$

$$\frac{A}{\omega_{\text{UE}}} \leq f_n \leq \frac{A}{\mu_{\min}}, \tag{6.27c}$$

$$h_n \leq f_n, \tag{6.27d}$$

$$R_{n,m} \geq R_{\min}, \tag{6.27e}$$

$$\tau \leq T, \tag{6.27f}$$

$$\tau = \max\{\frac{\omega_{\text{BS}}\omega_{\text{UE}}h_n f_n}{A^2} T_p\}, \tag{6.27g}$$

where $\boldsymbol{h} = [h_1, h_2]$ and $\boldsymbol{f} = [f_1, f_2]$. Problem (6.27) is still intractable with a non-convex objective. To further decompose the problem, we will iterate between the two clusters, i.e., we first fix Cluster 2 to optimize the beamwidths in Cluster 1 and then fix Cluster 1 to optimize the beamwidths in Cluster 2.

6.4.3 Optimal beamwidth search for each cluster

Let us assume the beamwidths of Cluster 2 are fixed and optimize the beamwidths in Cluster 1 as an example. The optimization for Cluster 2 is similar. In this case, the SINRs of $\text{UE}_{2,1}$ and $\text{UE}_{2,2}$ are fixed. We denote them by $\gamma_{2,1}^{\text{fix}}$ and $\gamma_{2,2}^{\text{fix}}$. We also denote the corresponding variables f_2 and h_2 as f_2^{fix} and h_2^{fix}. Then, the beamwidth

optimization problem for Cluster 1 is as follows:

$$\underset{h_1,\, f_1}{\text{maximize}} \quad \sum_{m=1}^{2} R_{1,m} + (1 - \frac{\tau}{T}) \sum_{m=1}^{2} \log_2(1 + \gamma_{2,m}^{\text{fix}}) \tag{6.28a}$$

$$\text{subject to} \qquad\qquad \frac{A}{\omega_{\text{BS}}} \le h_1 \le \frac{A}{\eta_{\min}}, \tag{6.28b}$$

$$\frac{A}{\omega_{\text{UE}}} \le f_1 \le \frac{A}{\mu_{\min}}, \tag{6.28c}$$

$$h_1 \le f_1, \tag{6.28d}$$

$$R_{1,m} \ge R_{\min}, \tag{6.28e}$$

$$(1 - \frac{\tau}{T}) \log_2(1 + \gamma_{2,m}^{\text{fix}}) \ge R_{\min}, \tag{6.28f}$$

$$\tau \le T, \tag{6.28g}$$

$$\tau = \max\{\frac{\omega_{\text{BS}}\omega_{\text{UE}}h_1 f_1}{A^2}T_p, \frac{\omega_{\text{BS}}\omega_{\text{UE}}h_2^{\text{fix}} f_2^{\text{fix}}}{A^2}T_p\}, \tag{6.28h}$$

To simplify Problem (6.28), we discuss how to pick the optimal value for τ and remove it from the objective function. There are two cases for the optimal τ:

- Case 1: $h_1 f_1 < h_2^{\text{fix}} f_2^{\text{fix}}$. In this case, the τ should be set to $\tau^* = \frac{\omega_{\text{BS}}\omega_{\text{UE}}h_2^{\text{fix}} f_2^{\text{fix}}}{A^2}T_p$. Then, the objective function should be $(1 - \frac{\tau^*}{T}) \sum_{n=1}^{2} \sum_{m=1}^{2} \log_2(1 + \gamma_{n,m})$.

- Case 2: $h_1 f_1 \ge h_2^{\text{fix}} f_2^{\text{fix}}$. In this case, the τ should be set according to the value of $h_1 f_1$, which is $\tau = \frac{\omega_{\text{BS}}\omega_{\text{UE}}h_1 f_1}{A^2}$. Then, the objective function should be $(1 - \frac{\omega_{\text{BS}}\omega_{\text{UE}}h_1 f_1 T_p}{A^2 T}) \sum_{n=1}^{2} \sum_{m=1}^{2} \log_2(1 + \gamma_{n,m})$.

Since the solution for the two cases are different and the objective function may change, the search for the optimal beamwidths is complicated and needs to be sim-

plified. To simplify, first, we introduce a variable $g_n = h_n f_n$. Then, the SINR $\text{UE}_{n,1}$ and $\text{UE}_{n,2}$ can be rewritten as

$$\gamma_{n,1} = \frac{P_{n,1}\beta_{n,1}^2(g_n + \xi^2)}{\sum_{\ell=1,q\neq n}^{2} P_{q,\ell}\beta_{n,1}^2(f_n + \xi)\xi + \sigma^2}, \tag{6.29}$$

$$\gamma_{n,2} = \frac{P_{n,2}\beta_{n,2}^2(g_n + \xi^2)}{P_{n,1}\beta_{n,2}^2(g_n + \xi^2) + \sum_{\ell=1,q\neq n}^{2} P_{q,\ell}\beta_{n,2}^2(f_n + \xi)\xi + \sigma^2}. \tag{6.30}$$

Since we fix the parameters for Cluster 2, $\log_2(1 + \gamma_{2,1}) + \log_2(1 + \gamma_{2,2})$ is a constant, which we denote by C. Then, we define a function $F(g_1, f_1) \triangleq \log_2(1 + \gamma_{1,1}) + \log_2(1 + \gamma_{1,2}) + C$ which has the following property:

Proposition 1. For $F(g_1, f_1)$ with its domain defined by $(f_1, g_1) \in [lb_f, ub_f] \times [lb_g, \min\{f_1^2, ub_g\}]$, $0 < lb_g \leq lb_f^2$ and $0 < ub_g \leq ub_f^2$, the maximum point lies on the boundary $g_1 = f_1^2$, $f_1 \in [lb_f, \sqrt{ub_g}]$.

Proof. See Appendix A. $\qquad\square$

Proposition 1 implies that if we want to find the maximum point of $F(g_1, f_1)$, we only need to search on the boundary $g_1 = f_1^2, f_1 \in [lb_f, ub_f]$. This simplifies $F(g_1, f_1)$ to $F(f_1^2, f_1)$. We further define $F_b(f_1) \triangleq F(f_1^2, f_1)$. Then, to find the maximum point of $F(g_1, f_1)$, we can perform a line search for $F_b(f_1)$ on $f_1 \in [lb_f, ub_f]$.

Next, we define the function $G(g_1, f_1) \triangleq (1 - \frac{\omega_{\text{BS}}\omega_{\text{UE}}g_1}{A^2}T_p)F(g_1, f_1)$. Function $G(g_1, f_1)$ has the following property:

Proposition 2. For $G(g_1, f_1)$ with its domain defined by $(f_1, g_1) \in [\sqrt{lb_g}, ub_f] \times [lb_g, \min\{f_1^2, ub_g\}]$, $ub_f, lb_g, ub_g > 0$, $ub_g < ub_f^2$, the maximum point lies on the boundary $g_1 = f_1^2, f_1 \in [\sqrt{lb_g}, \sqrt{ub_g}]$.
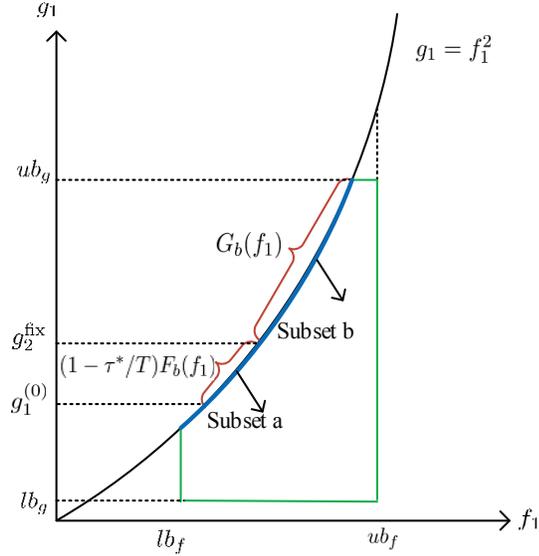
Figure 6.3: The search region division for Case 1.

Proof. See Appendix B. □

Proposition 2 implies that if we want to find the maximum point of $G(g_1,f_1)$, we only need to search on the boundary $g_1 = f_1^2, f_1 \in [lb_f, ub_f]$. This simplifies $G(g_1,f_1)$ into $G(f_1^2, f_1)$. We define $G_b(f_1) \triangleq G(f_1^2, f_1)$. Then, to find the maximum point of $G(g_1,f_1)$, we can perform the line search for $G_b(f_1)$ on $f_1 \in [lb_f, ub_f]$.

To find the maximum point for Problem (6.28), we plot its feasible region with boundaries colored in green and blue in Figs. 6.3 and 6.4. According to Propositions 1 and 2, the maximum point lies on the blue boundary and we only need to search on the blue boundary. However, the objective function varies along the blue boundary. To conduct an effective search, we need to divide the blue boundary into two different subsets. Moreover, different initial conditions lead to different division strategies. There are two cases:

- Case 1: $g_1^{(0)} < g_2^{\text{fix}}$, where $g_1^{(0)}$ is the initial point. In this case, along the
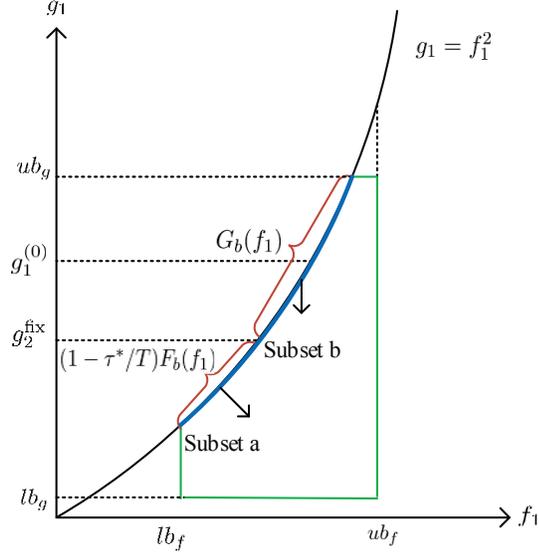
Figure 6.4: The search region division for Case 2.

blue boundary, when we increase g_1 from $g_1^{(0)}$ to g_2^{fix}, the objective function is $(1 - \frac{\tau^*}{T})F_b(f_1)$. If we continue to increase g_1, the objective function changes to $G_b(f_1)$. Then, the blue boundary is divided as shown in Fig. 6.3. On Subset a, we perform a line search over $(1 - \frac{\tau^*}{T})F_b(f_1)$ to find a maximum point $(f_1^{(F)}, (f_1^{(F)})^2)$. On Subset b, we perform a line search over $G_b(f_1)$ to find the maximum point $(f_1^{(G)}, (f_1^{(G)})^2)$. Then, we compare the values of $(1 - \frac{\tau^*}{T})F_b(f_1^{(F)})$ and $G_b(f_1^{(G)})$, to pick the larger one as the optimal solution (f_1^*, g_1^*).

- Case 2: $g_1^{(0)} \geq g_2^{\text{fix}}$, where $g_1^{(0)}$ is the initial point. In this case, when we decrease g_1 to be less than g_2^{fix}, the objective function changes from $G_b(f_1)$ to $(1 - \frac{\tau^*}{T})F_b(f_1)$. Then, the blue boundary is divided as shown in Fig. 6.4. On Subset a, we perform a line search over $(1 - \frac{\tau^*}{T})F_b(f_1)$ to find a maximum point $(f_1^{(F)}, (f_1^{(F)})^2)$. On Subset b, we perform a line search over $G_b(f_1)$ to find the maximum point $(f_1^{(G)}, (f_1^{(G)})^2)$. Then, we compare the values of $(1 - \frac{\tau^*}{T})F_b(f_1^{(F)})$

and $G_b(f_1^{(G)})$, to pick the larger one as the optimal solution (f_1^*, g_1^*).

While performing the line search, we also need to consider the minimum rate constraint. The details of line search for $(1 - \frac{\tau^*}{T})F_b(f_1)$ and $G_b(f_1)$ are described in Alg. 2 and Alg. 3, respectively.

Algorithm 2 Line search over $(1 - \frac{\tau^*}{T})F_b(f_1)$

1: Input τ^*, the search interval $[f_{\min}, f_{\max}]$ and the step size Δ;
2: Initialize $R_{\max} \quad 0$ and $f_1^{(F)} \quad f_{\min}$;
3: **for** $f_1 = f_{\min} : \Delta : f_{\max}$ **do**
4:     **if** $(1-\frac{\tau^*}{T})F_b(f_1) > R_{\max}$ and $(1-\frac{\tau^*}{T})\log_2(1+\gamma_{1,1}) \geq R_{\min}$ and $(1-\frac{\tau^*}{T})\log_2(1+\gamma_{1,2}) \geq$
    $R_{\min}$ **then**
5:         $R_{\max} \leftarrow (1 - \frac{\tau^*}{T})F_b(f_1)$;
6:         $f_1^{(F)} \quad f_1$;
7:     **end if**
8: **end for**
9: Return $R_{\max}$, $f_1^{(F)}$.

Algorithm 3 Line search over $G_b(f_1)$

1: Input the search interval $[f_{\min}, f_{\max}]$ and the step size Δ;
2: Initialize $R_{\max} \leftarrow 0$ and $f_1^{(G)} \leftarrow f_{\min}$;
3: **for** $f_1 = f_{\min} : \Delta : f_{\max}$ **do**
4:     $\tau \leftarrow \frac{\omega_{\mathrm{BS}}\omega_{\mathrm{UE}}f_1^2}{A^2}T_p$
5:     **if** $G_b(f_1) > R_{\max}$ and $(1-\frac{\tau}{T})\log_2(1+\gamma_{1,m}) \geq R_{\min}$ and $(1-\frac{\tau}{T})\log_2(1+\gamma_{2,m}^{\mathrm{fix}}) \geq R_{\min}$
    then
6:         $R_{\max} \quad G_b(f_1)$;
7:         $f_1^{(G)} \quad f_1$;
8:     **end if**
9: **end for**
10: Return $R_{\max}$, $f_1^{(G)}$.

6.4.4   Joint beamwidth optimization

For a fixed power allocation, our final beamwidth search algorithm, iterating between the beamwidth search for the two clusters, is presented in Alg. 4.

Algorithm 4 Joint Beamwidth Optimization

1: Set the sum-rate $R_{sum}[-1] \leftarrow 0$, the maximal iteration number $k_{max} \leftarrow 1000$ and the convergence threshold $\epsilon \leftarrow 10^{-3}$;

2: Input: $R_{sum}[0]$, $\mu_2, \eta_2, \mu_1^{(0)}, \eta_1(0)$, ω_{BS}, ω_{UE}, $\mu_{\min}$ and $\eta_{\min}$;

3: Calculate $f_1^{(0)}, g_1^{(0)}, f_2^{(0)}$ and $g_2^{(0)}$ according to (6.23), (6.24) and $g_n = f_n h_n$;

4: Calculate lb_f, ub_f, lb_g and ub_g based on ω_{BS}, ω_{UE}, $\mu_{\min}$ and $\eta_{\min}$;

5: $k \leftarrow 0$;

6: **while** $R_{sum}[k] - R_{sum}[k-1] \geq \epsilon R_{sum}[k-1]$ and $k \leq k_{max}$ **do**

7:   $k \leftarrow k+1$;

8:   **if** $g_1^{(k-1)} < g_2^{(k-1)}$ **then**

9:    $\tau^* \leftarrow \frac{\omega_{\mathrm{BS}} \omega_{\mathrm{UE}} g_2^{(k-1)} T_p}{A^2}$;

10:    Do line search for $(1 - \frac{\tau^*}{T}) F_b(f_1)$ on interval $f_1 \in [\sqrt{g_1^{(k-1)}}, \sqrt{g_2^{(k-1)}}]$ using Alg. 2 to get the maximum point $(f_1^{(F)}, (f_1^{(F)})^2)$;

11:    Do line search for $G_b(f_1)$ on interval $f_1 \in [\sqrt{g_2^{(k-1)}}, \sqrt{ub_g}]$ using Alg. 3 to get the maximum point $(f_1^{(G)}, (f_1^{(G)})^2)$;

12:    **if** $F_b(f_1^{(F)}) == 0$ and $G_b(f_1^{(G)}) == 0$ **then**

13:     $(f_1^{(k)}, g_1^{(k)}) \leftarrow (f_1^{(k-1)}, g_1^{(k-1)})$;

14:    **end if**

15:    Compare the value of $(1 - \frac{\tau^*}{T}) F_b(f_1^{(F)})$ and $G_b(f_1^{(G)})$, and pick the larger one as the optimal solution $(f_1^{(k)}, g_1^{(k)})$;

16:   **else if** $g_1^{(k-1)} \geq g_2^{(k-1)}$ **then**

17:    Do line search for $G_b(f_1)$ on interval $f_1 \in [\sqrt{g_2^{(k-1)}}, \sqrt{ub_g}]$ using Alg. 3 to get the maximum point $(f_1^{(G)}, (f_1^{(G)})^2)$;

18:    $\tau^* \leftarrow \frac{\omega_{\mathrm{BS}} \omega_{\mathrm{UE}} g_2^{(k-1)} T_p}{A^2}$;

19:    Do line search for $(1 - \frac{\tau^*}{T}) F_b(f_1)$ on interval $f_1 \in [lb_f, \sqrt{g_2^{(k-1)}}]$ using Alg. 2 to get the maximum point $(f_1^{(F)}, (f_1^{(F)})^2)$;

20:        **if** $F_b(f_1^{(F)}) == 0$ and $G_b(f_1^{(G)}) == 0$ **then**

21:           $(f_1^{(k)}, g_1^{(k)}) \quad (f_1^{(k-1)}, g_1^{(k-1)})$;

22:        **end if**

23:        Compare the value of $(1 - \frac{\tau^*}{T})F_b(f_1^{(F)})$ and $G_b(f_1^{(G)})$, and pick the larger one as the optimal solution $(f_1^{(k)}, g_1^{(k)})$;

24:    **end if**

25:    **if** $g_2^{(k-1)} < g_1^{(k)}$ **then**

26:        $\tau^* \quad \frac{\omega_{\mathrm{BS}}\omega_{\mathrm{UE}} g_1^{(k)} T_p}{A^2}$;

27:        Do line search for $(1 - \frac{\tau^*}{T})F_b(f_2)$ on interval $f_2 \in [\sqrt{g_2^{(k-1)}}, \sqrt{g_1^{(k)}}]$ using Alg. 2 to get the maximum point $(f_2^{(F)}, (f_2^{(F)})^2)$;

28:        Do line search for $G_b(f_2)$ on interval $f_2 \in [\sqrt{g_1^{(k)}}, \sqrt{ub_g}]$ using Alg. 3 to get the maximum point $(f_2^{(G)}, (f_2^{(G)})^2)$;

29:        **if** $F_b(f_2^{(F)}) == 0$ and $G_b(f_2^{(G)}) == 0$ **then**

30:           $(f_2^{(k)}, g_2^{(k)}) \quad (f_2^{(k-1)}, g_2^{(k-1)})$;

31:        **end if**

32:        Compare the value of $(1 - \frac{\tau^*}{T})F_b(f_2^{(F)})$ and $G_b(f_2^{(G)})$, and pick the larger one as the optimal solution $(f_2^{(k)}, g_2^{(k)})$;

33:    **else if** $g_2^{(k-1)} \geq g_1^{(k)}$ **then**

34:        Do line search for $G_b(f_2)$ on interval $f_2 \in [\sqrt{g_1^{(k)}}, \sqrt{ub_g}]$ using Alg. 3 to get the maximum point $(f_2^{(G)}, (f_2^{(G)})^2)$;

35:        $\tau^* \quad \frac{\omega_{\mathrm{BS}}\omega_{\mathrm{UE}} g_1^{(k)} T_p}{A^2}$;

36:        Do line search for $(1 - \frac{\tau^*}{T})F_b(f_2)$ on interval $f_1 \in [lb_f, \sqrt{g_1^{(k)}}]$ using Alg. 2 to get the maximum point $(f_2^{(F)}, (f_2^{(F)})^2)$;

37:        **if** $F_b(f_2^{(F)}) == 0$ and $G_b(f_2^{(G)}) == 0$ **then**

38:           $(f_2^{(k)}, g_2^{(k)}) \quad (f_2^{(k-1)}, g_2^{(k-1)})$;

39:        **end if**

40:　　　　Compare the value of $(1 - \frac{\tau^*}{T})F_b(f_2^{(F)})$ and $G_b(f_2^{(G)})$, and pick the larger one as the optimal solution $(f_2^{(k)}, g_2^{(k)})$;

41:　　**end if**

42:　　Calculate $R_{sum}[k]$;

43: **end while**

44: Calculate μ_n and η_n based on f_n, h_n, where $h_n = g_n/f_n$;

45: Return μ_n^* and η_n^*.

6.4.5　The joint algorithm

Based on Alg. 1 and Alg. 4, we can propose a joint optimization algorithm, which iterates between the power allocation and the beamwidth optimization. The details of the algorithm are described in Alg. 5.

Algorithm 5 Joint optimization

1: Set the sum-rate $R_{sum}[-1]$　　0, the maximal iteration number n_{max}　　1000 and the convergence threshold $\epsilon \leftarrow 10^{-3}$;
2: Choose a feasible start point $\mathbf{P}^*[0]$, $\boldsymbol{\mu}^*[0]$, $\boldsymbol{\eta}^*[0]$ and $\boldsymbol{\phi}^*[0]$;
3: $n \leftarrow 0$;
4: **while** $R_{sum}[n] - R_{sum}[n-1] \geq \epsilon R_{sum}[n-1]$ and $n \leq n_{max}$ **do**
5:　　$n \leftarrow n+1$;
6:　　Search the optimal beamwidth using Alg. 4 with $\mathbf{P}^*[n]$ to obtain $\boldsymbol{\mu}^*[n]$, $\boldsymbol{\eta}^*[n]$ and $\boldsymbol{\phi}^*[n]$;
7:　　Solve Problem (6.12) using Alg. 1 with $\boldsymbol{\mu}^*[n-1]$, $\boldsymbol{\eta}^*[n-1]$, $\boldsymbol{\phi}^*[n-1]$ to obtain $\mathbf{P}^*[n]$;
8: **end while**
9: Return $\mathbf{P}^*[n]$, $\boldsymbol{\mu}^*[n]$, $\boldsymbol{\eta}^*[n]$.

6.4.6　Convergence and complexity analysis

To prove the convergence of Alg. 5, we first need to prove the convergence of Alg. 1 and Alg. 4. The convergence of Alg. 1 has been proved in [130]. In Alg. 4, to maximize the sum-rate, we optimize the beamwidth for one cluster while keeping

the other cluster fixed. Such a step cannot decrease the sum-rate and generates a non-decreasing sequence of sum-rate values. Therefore, the convergence of Alg. 4 is guaranteed because the algorithm generates a sequence of non-decreasing sum-rates with an upper bound (the maximum sum-rate).

In Alg. 5, when allocating the power, we increase the sum-rate while keeping the beams in the feasible region. When optimizing the beamwidth, we search the feasible region for the beams to find the maximum sum-rate while guaranteeing the minimum rate constraint and keeping the powers in the feasible region. By doing so, we generate a monotonically increasing sequence with an upper bound (the maximum sum-rate), which proves the convergence.

Here, we provide the complexity analysis of the proposed algorithm. In our algorithm, we iteratively optimize the power allocation and beamwidth. In the power allocation algorithm, we use SPCA to gradually convexify the original non-convex problem. In each iteration, the complexity mainly lies in solving Problem (6.21). We use an off-the-shelf solution, i.e., CVX to solve Problem (6.21), which uses the interior-point method. The computational complexity of CVX is $\mathcal{O}((3M_{\mathrm{UE}})^{3.5})$, where M_{UE} is the total number of users and $3M_{\mathrm{UE}}$ is the number of variables in Problem (6.21). In the beamwidth optimization, we iteratively optimize the beamwidth for each sector. In each iteration, the main complexity lies in the line-search algorithm, with the complexity $\mathcal{O}(\frac{ub_f - lb_f}{\Delta})$, ub_f and lb_f are the upper bound and lower bound of variable f in Problem (6.28), respectively, and Δ is the stepsize of the line-search algorithm.

6.5 Simulation Results

In this section, we present the simulation results of the joint power and beamwidth optimization algorithm. Four UEs are considered which are divided in two clusters each with two UEs. It is assumed that the UEs inside each cluster have different distances from the BS. The closer UE to the BS is termed the near UE and the other UE is termed the far UE. Four multiple access techniques are investigated. The first technique is OMA in which UEs are served in different time slots. The second technique is a combination of OMA and NOMA called NOMA-OMA. In NOMA-OMA, UEs that belong to the same cluster are supported by a fixed-power NOMA and each cluster is supported by OMA at each time slot. The third technique is Fixed-NOMA in which all UEs are served by a fixed-power NOMA at one time slot. Finally, the fourth technique is the jointly optimized power and beamwidth NOMA system presented in Section 6.4, called Optimized-NOMA. For all techniques, first, the beams are trained and then the data transmission is done.

To evaluate the performance of the Optimized-NOMA, the parameters are set as follows. The minimum rate for all UEs is assumed to be $R_{\min} = 0.1$ bits/s/Hz. Further, for the Fixed-NOMA, we allocate $\frac{1}{5}$ of the total power to the near UE and $\frac{4}{5}$ of the total power to the far UE as done in [58]. Also, the power is equally divided between the two clusters. The SNR used in the simulations indicates the transmit SNR, i.e., SNR=$\frac{P_{\text{tot}}}{\sigma^2}$, $\sigma^2 = 1$. In the first cluster, the channel gains of the near and far UEs from the BS are $\beta_{1,1}^2 = -17$dB and $\beta_{1,2}^2 = -26.5$dB. In the second cluster, the channel gains of the near and far UEs are $\beta_{2,1}^2 = -19$dB and $\beta_{2,2}^2 = -25$dB. Also, the side lobe level is constant and is given as $\xi = 0.1$. For the Optimized-NOMA, we use $\omega_{\text{BS}} = \omega_{\text{UE}} = 120°$ and $\eta_{\min} = \mu_{\min} = 5°$ unless it is mentioned otherwise.
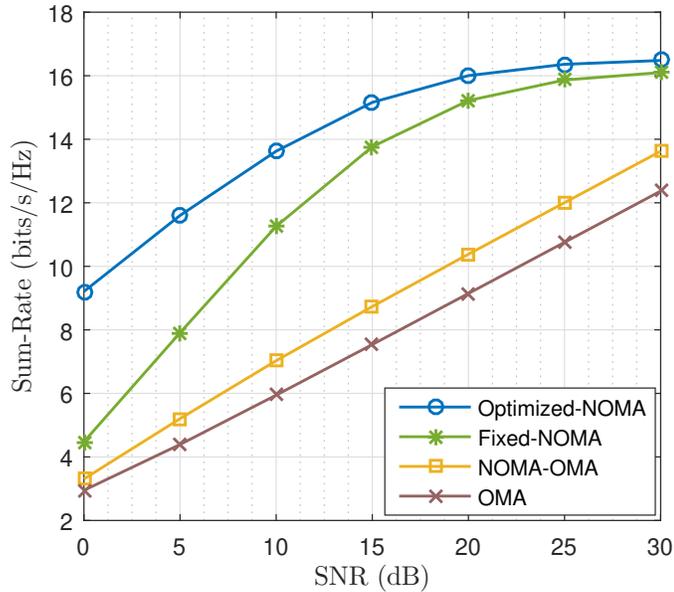
Figure 6.5 **Performance of the sum-rate versus SNR for a large channel coher-ence time, i.e., $T = 5 \times 10^3 T_p$.**

Further, the convergence threshold is set to $\epsilon = 10^{-3}$.

Fig. 6.5 demonstrates the performance of the sum-rate versus SNR. It is assumed that $T = 5 \times 10^3 T_p$ which indicates a large channel coherence time and $\eta = \mu_1 = \mu_2 = 10°$. For all SNRs, by increasing SNR the sum-rate increases. The Optimized-NOMA achieves the highest sum-rate. Especially, at low SNRs, the performance gap is larger. For instance, at SNR $= 0$dB the gap between the Optimized-NOMA and Fixed-NOMA is more than 5 bits/s/Hz which reveals our joint optimization algorithm designs the powers and beamwidths very efficiently. As SNR increases, the gap decreases which is due to the fact that in the Fixed-NOMA the BS and UEs steer strong beams even if the powers and beamwidths are not optimized. The Fixed-NOMA technique preforms better than the NOMA-OMA and OMA techniques. The reason is that the Fixed-NOMA serves all the users at the same time and takes the
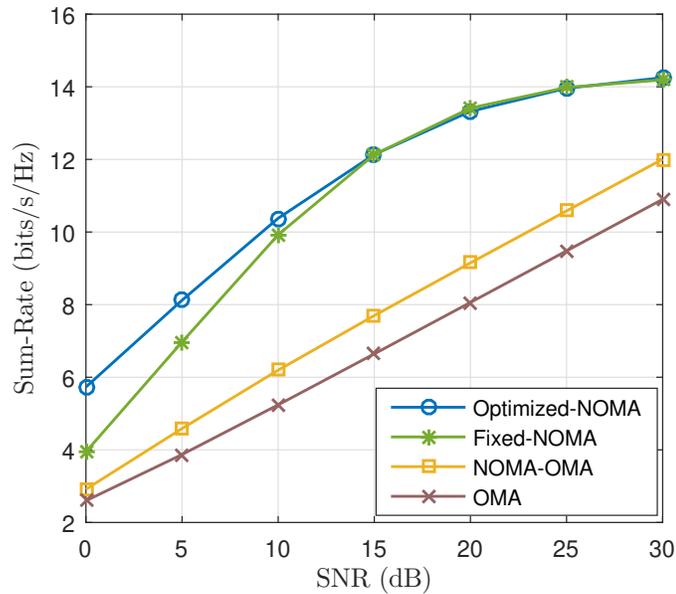
Figure 6.6   **Performance of the sum-rate versus SNR for a short channel coher-ence time** $T = 1 \times 10^3 T_p$.

advantages of the spectrum sharing among UEs.

In Fig. 6.6, we repeat the same simulation as in Fig. 6.5 for a relatively short channel coherence time, i.e., $T = 1 \times 10^3 T_p$. Similarly, by increasing SNR, the sum-rate increases for all the techniques. However, compared to Fig. 6.5, at low SNRs, the rate gap between the Optimized-NOMA and Fixed-NOMA is small. Moreover, at high SNR regions, these two techniques achieve identical sum-rates. This is because when the channel coherence time is short, the optimization algorithm does not allocate a large portion of T to the beam-alignment, e.g., τ is small. Thus, the optimized beamwidths are not narrow enough to provide higher gain. Also, at high SNRs, the optimized powers have trivial effects on the sum-rate compared to the predefined fixed values. For a smaller coherence time, the optimized-NOMA shows better performance only at low SNRs. Nevertheless, severe path loss and shadowing in mmWave bands

makes the low SNR regime very crucial. Especially, NOMA is supposed to consider near and far users where the far users likely receive signal through NLoS low SNR channels [67]. We emphasize that at high SNRs, by increasing the coherence time, the rate gap between the optimized-NOMA and fixed-NOMA becomes larger (See Fig. 6.5).

Fig. 6.7 shows the sum-rate performance versus the normalized channel coherence time, i.e., T/T_p for the moderate SNR= 20dB. In this simulation, two sets of beamwidths are considered for the first three techniques: (i) $\eta = \mu = 5°$ (narrow beamwidth) and (ii) $\eta = \mu = 10°$ (relatively wide beamwidth). The Optimized-NOMA outperforms the other techniques for both sets of the beamwidths. For the Fixed-NOMA with $\eta = \mu = 10°$ and short normalized channel coherence times, the performance is very close to that of the Optimized-NOMA. This is expected as we explained before. However, for $\eta = \mu = 5°$ and short normalized channel coherence times, the sum-rate of the Fixed-NOMA is much smaller than that of the Optimized-NOMA. This is because at small T/T_p, the Fixed-NOMA assigns more time to the beam-alignment and leaves less time for the data transmission. As the normalized time goes up, more time is available for data transmission, and the narrow beam provides a higher sum-rate. This statement also is supported by Fig. 6.7 where at large channel coherence times, the Optimized-NOMA selects the minimum beamwidth. Hence, the Optimized-NOMA and the Fixed-NOMA with $\eta = \mu = 5°$ achieve identical sum-rates at large normalized channel coherence times. Using Optimized-NOMA, a wide beamwidth is preferred for short T/T_p while a narrow beamwidth is preferred for large T/T_p. In Fig. 6.8, we simulate the performance of the sum-rate versus the minimum beamwidth of BS and UEs. The simulation is done for SNRs 10dB, 20dB,
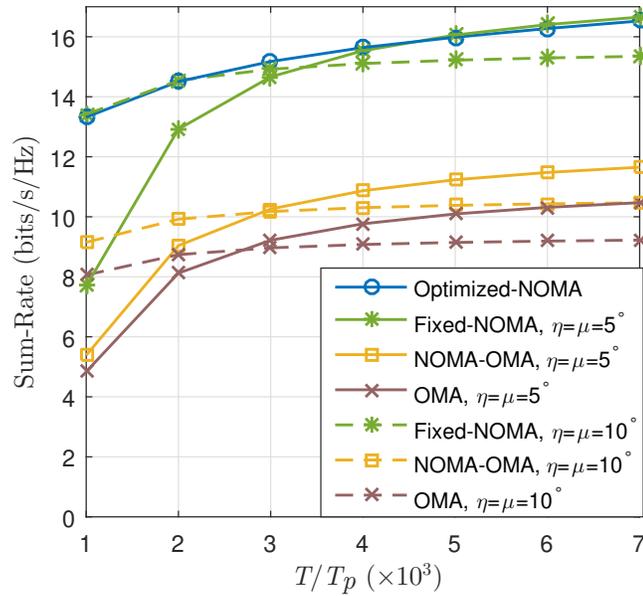
Figure 6.7 Performance of the sum-rate versus T/T_p for SNR=20dB.
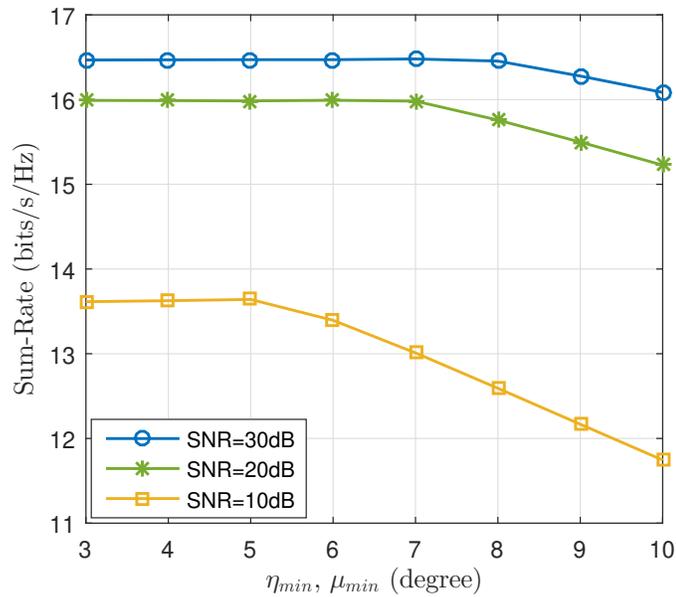
Figure 6.8 Performance of the sum-rate of the Optimized-NOMA versus minimum BS and UE beamwidth ($\eta_{\mathbf{min}}$ and $\mu_{\mathbf{min}}$) for $T = 5 \times 10^3 T_p$ and various SNRs.

and 30dB and $T = 5 \times 10^3 T_p$. In practice, the number of antennas at the BS and UEs is limited and even for large T/T_p, a narrow beamwidth may not be generated. At high SNR, i.e., 30dB, increasing the minimum beamwidth does not affect the sum-rate severely. As such, compared to $\eta_{\min} = \mu_{\min} = 3°$, at 10°, the sum-rate is reduced only by 0.5 bits/s/Hz. At SNR= 20dB, the sum-rate drops about 0.8 bits/s/Hz which is larger than the drop at SNR= 30dB. The minimum beamwidth has a major effect at SNR= 10dB. When SNR is low, narrow beams can still provide high gains to compensate for the low SNR. As the minimum beamwidth increases and the SNR is low, the optimization algorithm cannot select narrow beams. As a result, the sum-rate dramatically decreases. In this case, simulation results indicate that decreasing the beamwidth from 3° to 10° decreases the sum-rate by about 2 bits/s/Hz.

6.6   Summary and Conclusions

NOMA is incorporated into mmWave hybrid beamforming systems. We also consider the beam-alignment time because of the limited channel coherence time in mmWave directional communications. By combining the exhaustive search and tone-based beam-alignment algorithms, a new beam-alignment algorithm is employed. The formulated sum-rate expression consists of the channel coherence time and beam-alignment time. To maximize the sum-rate, a joint power allocation and beamwidth control optimization problem is solved by an algorithm which iterates between the power allocation and the beamwidth optimization. The non-convex power allocation is solved by the log-reformulation and SPCA. The beamwidth optimization is solved by iterating between the two clusters. A boundary-search algorithm is proposed to reduce the search complexity for the beamwidth in each cluster. The numerical results demonstrate that an efficient power allocation and beam-alignment time can lead to

higher sum-rates compared to the conventional mmWave-NOMA without optimized parameters, NOMA-OMA, and OMA. Also, at low SNRs, the size of the antenna array is a major obstacle in achieving higher sum-rates.

CHAPTER 7:

CONCLUSION AND FUTURE WORK

In this chapter, we discuss the summary of the results and conclude each chapter. Also, we talk about the future work directions.

7.1   Summary and Conclusion

In this dissertation, we have focused on millimeter-Wave (mmWave) bands communication. In mmWave bands, NLoS channels are vulnerable to severe shadowing and high path loss. Thus, these channels may experience frequent outage which leads to an unreliable communication. Further, in LoS channels due to ill-conditioned channels achieving multiplexing gain is very challenging. As the distance between transmitter and receiver becomes larger, the problem becomes worse. Further, due to the directional transmission, mmWave communication is susceptible to beam misalignment which can dramatically impact the sum-rate. Moreover, the directional transmission leads to a limited channel coherence time when users move. This causes a portion of the transmission frame to be allocated to beam-alignment which leads to a trade-off between the beamwidth resolution and the sum-rate.

To this end, in Chapter 3, we designed a new multiple-input multiple-output (MIMO) system using lens-based reconfigurable antennas. The system consists of multiple separated lens antennas at the transceiver. Further, thanks to their reconfig-

urability, the channel matrix can be reconfigured. Such that, by adjusting the distance between the antennas and also using the reconfigured channels, the ill-conditioned channel issue can be resolved. Further, in mmWave bands, each scatterer in NLoS channels experiences an independent shadowing. Taking this important property into account and using the proposed MIMO along with designing suitable space-time block codes, the full-diversity gain is achieved. Our simulation results indicated that the proposed MIMO can remarkably reduce the error and outage probability in NLoS channels.

In Chapter 4, we designed two new multiple access techniques based on the reconfigurable lens antenna. In the first design which is called reconfigurable antenna multiple access (RAMA), we aimed to transmitted only the intended signal for every user to avoid the inter-user interference. Our rate analysis revealed that, the proposed RAMA outperforms non-orthogonal multiple access (NOMA) in terms of the sum-rate. Since RAMA is only applicable for users with distinguished directions, in the second technique, we combined RAMA with NOMA, i.e., RA-NOMA. Such that, the users with similar channel gain but different directions are served via RAMA. The users with similar directions and different channel gains are served via NOMA. The most important observation is that in the RA-NOMA the number of radio frequency (RF) chains can be smaller than the number clusters. For this case, our simulation results showed that the RA-NOMA gives higher sum-rate than NOMA.

In Chapter 5, we evaluated the impact of beam misalignment in mmWave-NOMA systems. To this end, we first designed the unknown parameters to maximize the sum-rate for the case of without beam misalignment and derived a lower bound on the achievable rate. Next, we modeled the misaligned effective channel vector with

regard to the effective channel vector of the corresponding cluster-head and the other clusters. Using the model, we derived a lower bound on the achievable rate. Finally, we found the rate gap between the aligned and misaligned effective channel vectors. The simulation results revealed that the obtained bounds are tight. Also, we observed that the users with higher channel gain are impacted by beam misalignment more severe than the users with low channel gain. This is because, in the high channel gain users, the intra-cluster interference is not large. Hence, the presence of inter-cluster interference has a serious impact.

In Chapter 6, we focused on joint beamwidth design and power allocation in mmWave-NOMA systems with a limited channel coherence time. Due to the limited channel coherence time, users' direction information are valid only for a short time. Therefore, at the beginning of each transmission frame, the direction should be estimated which causes rate loss. In mmWave-NOMA systems, because of the existence of low channel gain users, the problem becomes crucial. If a short time is spent on the direction estimation, the low channel user and the BS may not be able found the best beam pair. On the hand, spending more time on the beam-alignment brings about allocating less time for data transmission. In order to find the optimal value for the beam-alignment, we formulated an optimization problem. Then, by iteration between the power allocation and beamwidth optimization phases, we solved the problem. The simulation results showed that the optimized mmWave-NOMA can significantly improve the sum-rate performance especially when the channel coherence time is fairly high.

7.2   Future Work

The research accomplished in this dissertation can open new research directions highlighted in the following:

- In Chapter 3, we considered ideal lens and beam switching network. In practice, neither lens nor beam switching network is ideal. The non-ideal lens can impose interference on the signal transmitted by the other lens. This situation may degrade the full-diversity gain. the same issue may cause by non-ideal switches. In [132], the impact of spillover losses due to the switching and Rotman lens is studied. Similar work can be done for the proposed RA-MIMO system.

- Beam switching network and lens imperfections can also affect the sum-rate of the RAMA and RA-NOMA proposed in Chapter 4. As a future work, this important issue needs more investigation. On the other hand, the proposed RAMA and RA-NOMA required phase-shifters that are as fast as the symbol-rate. In mmWave bands, this assumption may be practical or may cause additional hardware costs with the current technology. New efficient solutions are needed. Further, lens operates as a passive beamformer in which the beamwidth is not adjustable in RAMA. Integration of RAMA with the emerging technologies such as reconfigurable intelligent surfaces can make RAMA able to perform passive-active beamforming.

- In Chapter 5, we observed that the beam misalignment reduces the sum-rate. As a future work, a robust user clustering and beamforming design that takes the beam misalignment into account is required. Also, in Chapter 6, most of the current beam-alignment methods are not compatible with mmWave-NOMA sys-

tems. Therefore, a new beam-alignment method seems necessary in mmWave-NOMA systems with limited channel coherence time such that the method takes the channel gain difference between NOMA users into account.

BIBLIOGRAPHY

[1] M. A. Almasi, H. Mehrpouyan, V. Vakilian, N. Behdad, and H. Jafarkhani, "Reconfigurable antennas in mmWave MIMO systems," *arXiv preprint arXiv:1710.05111*, 2017.

[2] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Introduction to ultra reliable and low latency communications in 5G," *arXiv preprint arXiv:1704.05565*, 2017.

[3] J. Guo, S. Karimi-Bidhendi, and H. Jafarkhani, "Energy-efficient node deployment in wireless Ad-hoc sensor networks," *in Proc. IEEE International Conference on Communications (ICC)*, pp. 1–6, June 2020.

[4] S. Karimi-Bidhendi, J. Guo, and H. Jafarkhani, "Energy-efficient node deployment in heterogeneous two-tier wireless sensor networks with limited communication range," *IEEE Trans. Wireless Commun.*, pp. 1–1, Sept. 2020.

[5] ——, "Using quantization to deploy heterogeneous nodes in two-tier wireless sensor networks," *in Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 1502–1506, July 2019.

[6] M. Series, "IMT vision–framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU*, vol. 2083, Sept. 2015.

[7] D. T. Emerson, "The work of jagadis chandra bose: 100 years of millimeter-wave research," *IEEE Trans. Microw. Theory Techn.*, vol. 45, no. 12, pp. 2267–2273, Dec. 1997.

[8] R. E. Ziemer, "An overview of millimeter wave communications," *in Proc. IEEE 14th European Microwave Conference*, pp. 3–8, 1984.

[9] H. H. Hmimy and S. C. Gupta, "Performance of frequency-hopped npcsma for broad-band personal communication services (B-PCS) at millimeter waves in an urban mobile radio environment," *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 90–97, Jan. 1999.

[10] Hao Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, April 2002.

[11] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sept. 2017.

[12] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond.* Springer, 2018.

[13] M. A. Almasi, H. Mehrpouyan, V. Vakilian, N. Behdad, and H. Jafarkhani, "A new reconfigurable antenna MIMO architecture for mmWave communication," *in Proc. IEEE ICC*, pp. 1–7, May 2018.

[14] M. A. Almasi and H. Mehrpouyan, "Reconfigurable antenna multiple access for millimeter wave systems," Nov. 21 2019, uS Patent App. 16/209,350.

[15] M. A. Almasi, R. Amiri, H. Jafarkhani, and H. Mehrpouyan, "mmWave lens-based MIMO system for suppressing small-scale fading and shadowing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5292–5306, 2020.

[16] M. A. Almasi, H. Mehrpouyan, D. Matolak, C. Pan, and M. Elkashlan, "Reconfigurable antenna multiple access for 5G mmWave systems," *in Proc. IEEE ICC Workshops*, pp. 1–6, May 2018.

[17] M. A. Almasi, R. Amiri, M. Vaezi, and H. Mehrpouyan, "Lens-based millimeter wave reconfigurable antenna NOMA," *in Proc. IEEE ICC Workshops*, pp. 1–5, 2019.

[18] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[19] M. A. Almasi and H. Mehrpouyan, "Non-orthogonal multiple access based on hybrid beamforming for mmwave systems," *in Proc. IEEE Veh. Technol. Conf., Fall*, Aug. 2018.

[20] M. A. Almasi, M. Vaezi, and H. Mehrpouyan, "Impact of beam misalignment on hybrid beamforming NOMA for mmWave communications," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 1–13, Mar. 2019.

[21] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[22] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.

[23] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[24] M. A. Almasi, L. Jiang, H. Jafarkhani, and H. Mehrpouyan, "Joint beamwidth and power optimization in mmWave hybrid beamforming-NOMA systems," *Submitted to IEEE Trans. Wireless Commun.*

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

[26] T. S. Rappaport *et al.*, *Millimeter wave wireless communications.* Pearson Education, 2014.

[27] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[28] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sept. 2017.

[29] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[30] J. Kim and I. Lee, "802.11 WLAN: history and new enabling MIMO techniques for next generation standards," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 134–140, Mar. 2015.

[31] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[32] L. Jiang and H. Jafarkhani, "Multi-user analog beamforming in millimeter wave MIMO systems based on path angle information," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 608–619, Jan. 2019.

[33] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[34] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[35] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[36] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, Nov. 2016.

[37] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.

[38] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1194–1206, June 2014.

[39] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[40] T. Kwon, Y.-G. Lim, B.-W. Min, and C.-B. Chae, "RF lens-embedded massive MIMO systems: Fabrication issues and codebook design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2256–2271, July 2016.

[41] Y. Zeng and R. Zhang, "Cost-effective millimeter-wave communications with lens antenna array," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 81–87, Aug. 2017.

[42] B. He and H. Jafarkhani, "Low-complexity reconfigurable MIMO for millimeter wave communications," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5278–5291, Nov. 2018.

[43] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," *in Proc. IEEE GLOBECOM*, pp. 3679–3684, Dec. 2013.

[44] B. Badamchi *et al.*, "Design of compact reconfigurable ultra-wideband slot antenna with switchable single/dual band notch functions," *IET Microw., Antennas & Propag.*, vol. 8, no. 8, pp. 541–548, June 2014.

[45] B. Badamchi, A. Valizade, P. Rezaei, and Z. Badamchi, "A reconfigurable square slot antenna with switchable single band, UWB and UWB with band notch function performances," *ACESJ*, vol. 29, no. 5, pp. 383–390, May 2014.

[46] P.-Y. Qin, A. R. Weily, Y. J. Guo, and C.-H. Liang, "Polarization reconfigurable U-slot patch antenna," *IEEE Trans. Antennas Propag.*, vol. 58, no. 10, pp. 3383–3388, July 2010.

[47] H. Yazdani and A. Vosoughi, "On the spectrum sensing, beam selection and power allocation in cognitive radio networks using reconfigurable antennas," *in Proc. IEEE CISS*, pp. 1–7, Mar. 2019.

[48] B. A. Cetiner, H. Jafarkhani, Jiang-Yuan Qian, Hui Jae Yoo, A. Grau, and F. De Flaviis, "Multifunctional reconfigurable MEMS integrated antennas for adaptive MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 62–70, Dec. 2004.

[49] A. Grau, H. Jafarkhani, and F. De Flaviis, "A reconfigurable multiple-input multiple-output communication system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1719–1733, May 2008.

[50] F. Fazel, A. Grau, H. Jafarkhani, and F. De Flaviis, "Space-time-state block coded MIMO communication systems using reconfigurable antennas," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6019–6029, Dec. 2009.

[51] V. Vakilian, J.-F. Frigon, and S. Roy, "Space-frequency block code for MIMO-OFDM communication systems with reconfigurable antennas," *in Proc. IEEE GLOBECOM*, pp. 4221–4225, Dec. 2013.

[52] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[53] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," *in Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, pp. 611–615, Sep. 2013.

[54] Y. Saito *et al.*, "Non-orthogonal multiple access (NOMA) for cellular future radio access," *in Proc. IEEE Veh. Technol. Conf., Spring*, pp. 1–5, Jun. 2013.

[55] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[56] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, Jun. 2015.

[57] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[58] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[59] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[60] ——, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.

[61] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge university press, 2005.

[62] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam NOMA for hybrid mmwave systems," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1705–1719, Feb. 2019.

[63] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[64] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, "System-level performance of downlink NOMA for future LTE enhancements," *in Proc. IEEE Globecom Workshops*, pp. 66–70, Dec. 2013.

[65] S. Niknam and B. Natarajan, "On the regimes in millimeter wave networks: Noise-limited or interference-limited?" *in Proc. IEEE ICC Workshops*, pp. 1–6, May 2018.

[66] R. J. Weiler, W. Keusgen, A. Maltsev, T. Kühne, A. Pudeyev, L. Xian, J. Kim, and M. Peter, "Millimeter-wave outdoor access shadowing mitigation using beamforming arrays," *in Proc IEEE Eur. Conf. Antennas Propag.*, pp. 1–5, Apr. 2016.

[67] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[68] M. Flament and M. Unbehaun, "Impact of shadow fading in a mm-wave band wireless network," *In Proc. IEEE WPMC*, pp. 427–432, Nov. 2000.

[69] J. Choi, "On the macro diversity with multiple BSs to mitigate blockage in millimeter-wave communications," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1653–1656, Sept. 2014.

[70] A. K. Gupta, J. G. Andrews, and R. W. Heath, "Macrodiversity in cellular networks with random blockages," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 996–1010, Feb. 2018.

[71] H. Jafarkhani, *Space-Time Coding: Theory and Practice.* Cambridge University Press, 2005.

[72] R. J. Weiler, M. Peter, W. Keusgen, and M. Wisotzki, "Measuring the busy urban 60 GHz outdoor access radio channel," *In Proc. IEEE ICUWB*, pp. 166–170, Sept. 2014.

[73] V. Vakilian *et al.*, "System for resolving channel sparsity in multiple-input multiple-output systems using (re) configurable antennas, and methods of making and using the same," Oct. 3 2016, US Patent App. 15/284,123.

[74] B. Schoenlinner, X. Wu, J. P. Ebling, G. V. Eleftheriades, and G. M. Rebeiz, "Wide-scan spherical-lens antennas for automotive radars," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 9, pp. 2166–2175, Sept. 2002.

[75] M. D. Wright, W. Baron, J. Miller, J. Tuss, D. Zeppettella, and M. Ali, "MEMS reconfigurable broadband patch antenna for conformal applications," *IEEE Trans. Antennas Propag.*, vol. 66, no. 6, pp. 2770–2778, June 2018.

[76] Fernandes *et al.*, "Dielectric lens antennas," *Handbook of antenna technologies*, pp. 1–54, 2014.

[77] M. A. B. Abbasi, V. F. Fusco, H. Tataria, and M. Matthaiou, "Lens-based beamformer for low-complexity millimeter-wave cellular systems," *in Proc. ESA Workshop on Millimetre-Wave Techn. and Applications*, pp. 1–4, Oct. 2018.

[78] Z. Gua, X. Wang, and W. Heng, "Millimeter-Wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384 – 5394, Aug. 2017.

[79] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[80] M. A. Almasi, R. Amiri, and H. Mehrpouyan, "A new millimeter wave MIMO system for 5G networks," *arXiv preprint arXiv:1807.04851*, 2018.

[81] F. Hansen and F. I. Meno, "Mobile fading—Rayleigh and lognormal superimposed," *IEEE Trans. Veh. Technol.*, vol. 26, no. 4, pp. 332–335, Nov. 1977.

[82] A. Abdi and M. Kaveh, "K distribution: an appropriate substitute for Rayleigh-lognormal distribution in fading-shadowing wireless channels," *Electron. Lett.*, vol. 34, no. 9, pp. 851–852, Apr. 1998.

[83] U. Karabulut *et al.*, "Spatial and temporal channel characteristics of 5G 3D channel model with beamforming for user mobility investigations," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 38–45, Dec. 2018.

[84] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005.

[85] S. Bryan, G. Che, C. Groppi, P. Mauskopf, and M. Underhill, "A compact filter-bank waveguide spectrometer for millimeter wavelengths," *IEEE Trans. THz Sci. Technol.*, vol. 5, no. 4, pp. 598–604, July 2015.

[86] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[87] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[88] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[89] Y. Shang and X.-G. Xia, "Space–time block codes achieving full diversity with linear receivers," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4528–4547, Oct. 2008.

[90] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1420–1425, Apr. 2007.

[91] S. L. H. Nguyen, K. Haneda, J. Jarvelainen, A. Karttunen, and J. Putkonen, "On the mutual orthogonality of millimeter-Wave massive MIMO channels," *in Proc. IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2015.

[92] Z. Ding, L. Dai, R. Schober, and H. V. Poor, "NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks," *IEEE Commun. Lett.*, Aug. 2017.

[93] B. Wang, L. Dai, Z. Wang, N. Ge, and S. Zhou, "Spectrum and energy-efficient beamspace MIMO-NOMA for millimeter-wave communications using lens antenna array," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2370–2382, Oct. 2017.

[94] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.

[95] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb 2016.

[96] D. B. Cline and G. Samorodnitsky, "Subexponentiality of the product of independent random variables," *Stochastic Process. Appl.*, vol. 49, no. 1, pp. 75–98, Jan 1994.

[97] Y.-L. Wu, H. Zhou, Y.-X. Zhang, and Y.-A. Liu, "An unequal wilkinson power divider for a frequency and its first harmonic," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 11, pp. 737–739, Nov. 2008.

[98] X. Ge, S. Tu, G. Mao, C. Wang, and T. Han, "5G ultra-dense cellular networks," *IEEE Commun. Mag.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.

[99] J. Choi, "NOMA: Principles and recent results," *in Proc. IEEE ISWCS*, pp. 349–354, Aug. 2017.

[100] Z. Yang, W. Xu, C. Pan, Y. Pan, and M. Chen, "On the optimality of power allocation for NOMA downlinks with individual QoS constraints," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1649–1652, Jul. 2017.

[101] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge University Press, 2004.

[102] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access*, vol. 5, pp. 7667–7681, Feb. 2017.

[103] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X.-G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter-wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May 2018.

[104] W. Wu and D. Liu, "Non-orthogonal multiple access based hybrid beamforming in 5G mmWave systems," *in Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, pp. 1–7, Oct. 2017.

[105] Y. Zhou, V. W. S. Wong, and R. Schober, "Coverage and rate analysis of millimeter wave NOMA networks with beam misalignment," *IEEE Trans. Wireless Commun.*, pp. 1–1, Oct. 2018.

[106] J. Wildman, P. H. J. Nardelli, M. Latva-aho, and S. Weber, "On the joint impact of beamwidth and orientation error on throughput in directional wireless poisson networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 7072–7085, Dec. 2014.

[107] M. Min, D. Kim, H. Kim, and G. Im, "Opportunistic two-stage feedback and scheduling for MIMO downlink systems," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 312–324, Jan. 2013.

[108] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.

[109] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *arXiv preprint arXiv:1809.07224*, 2018.

[110] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully non-orthogonal communication for massive access," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1717–1731, Apr. 2018.

[111] Q. Zhang, Q. Li, and J. Qin, "Robust beamforming for nonorthogonal multiple-access systems in MISO channels," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10 231–10 236, Dec. 2016.

[112] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[113] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[114] X. Chen, Z. Zhang, C. Zhong, and D. W. K. Ng, "Exploiting multiple-antenna techniques for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2207–2220, Oct. 2017.

[115] R. Strichartz, *The Way of Analysis*, ser. Jones and Bartlett books in mathematics. Jones and Bartlett Publishers, 2000.

[116] H. Shokri-Ghadikolaei, L. Gkatzikis, and C. Fischione, "Beam-searching and transmission scheduling in millimeter wave communications," *in Proc. IEEE ICC*, pp. 1292–1297, June 2015.

[117] J. Palacios, D. De Donno, D. Giustiniano, and J. Widmer, "Speeding up mmwave beam training through low-complexity hybrid transceivers," *in Proc. IEEE 27th PIMRC*, pp. 1–7, Sep. 2016.

[118] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[119] N. Eshraghi, V. Shah-Mansouri, and B. Maham, "Fair beamwidth selection and resource allocation for indoor millimeter-wave networks," *in Proc. IEEE ICC*, pp. 1–6, May 2017.

[120] C. Pradhan, H. Chen, Y. Li, and B. Vucetic, "Joint beamwidth and energy optimization for multi-user millimeter wave communications," *in Proc. IEEE ICC Workshops*, pp. 1–6, May 2018.

[121] R. A. Hassan *et al.*, "Multi-user beam-alignment for millimeter-wave networks," *in Proc. ITA*, pp. 1–7, Feb. 2018.

[122] M. Hussain and N. Michelusi, "Energy-efficient interactive beam alignment for millimeter-wave networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 838–851, Feb. 2019.

[123] W. Hao, F. Zhou, Z. Chu, P. Xiao, R. Tafazolli, and N. Al-Dhahir, "Beam alignment for MIMO-NOMA millimeter wave communication systems," *in Proc. IEEE ICC*, pp. 1–6, May 2019.

[124] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, 2013.

[125] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: Large deviations analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1619–1631, Jul. 2017.

[126] D. De Donno, J. Palacios, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3646–3660, June 2017.

[127] C. A. Balanis, *Antenna theory: analysis and design.* John wiley & sons, 2016.

[128] 3GPP R1-154999, "TP for classification of MUST schemes," *TSG-RAN WG1 #82, Beijing, China*, Aug. 24-28, 2015.

[129] W.-C. Li, T.-H. Chang, C. Lin, and C.-Y. Chi, "Coordinated beamforming for multiuser MISO interference channel under rate outage constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1087–1103, Mar. 2012.

[130] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *Journal of Global Optimization*, vol. 47, no. 1, pp. 29–51, May 2010.

[131] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[132] M. A. Babar Abbasi, H. Tataria, V. F. Fusco, and M. Matthaiou, "On the impact of spillover losses in 28 GHz rotman lens arrays for 5G applications," *IEEE MTT-S IMWS-5G*, pp. 1–3, 2018.

[133] M. K. Murugan and S. S. Adams, "Hurwitz-Radon inspired maximal three-dimensional real orthogonal designs," *Australasian Journal of Combinatorics*, vol. 54, pp. 151–162, Oct. 2012.

[134] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2012.

[135] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[136] D. A. Harville, *Matrix algebra from a statistician's perspective*. Taylor & Francis Group, 1998.

[137] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with log-normal components," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1441–1462, Sept. 1982.

[138] D. Schleher, "Generalized Gram-Charlier series with application to the sum of log-normal variates (corresp.)," *IEEE Trans. Inf. Theory*, vol. 23, no. 2, pp. 275–280, Mar. 1977.

[139] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Comparison of methods of computing lognormal sum distributions and outages for digital wireless applications," *in Proc. IEEE ICC*, pp. 1270–1275 vol.3, May 1994.

[140] N. C. Beaulieu and Qiong Xie, "Minimax approximation to lognormal sum distributions," *in Proc. IEEE VTC*, vol. 2, pp. 1061–1065 vol.2, Apr. 2003.

[141] L. Zhao and J. Ding, "Least squares approximations to lognormal sum distributions," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 991–997, Mar. 2007.

[142] X. Li, Z. Wu, V. D. Chakravarthy, and Z. Wu, "A low-complexity approximation to lognormal sum distributions via transformed log skew normal distribution," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 4040–4045, Oct. 2011.

[143] K. Scharnhorst, "Angles in complex vector spaces," *Acta Applicandae Mathematica*, vol. 69, no. 1, pp. 95–103, 2001.

[144] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[145] P. A. Dighe, R. K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 694–703, Apr. 2003.

APPENDIX A:

PROOFS OF CHAPTER 2

A.1   Proof of Lemma 1

By plugging (3.9) into (3.5), the received signal can be rewritten as

$$
\begin{aligned}
\mathbf{Y} &= \sqrt{\frac{P_t}{N_t N_r}} \big(\mathbf{A}_1 s_1 + \mathbf{A}_2 s_2 + \cdots + \mathbf{A}_L s_L \big) \mathbf{H}_g + \mathbf{Z} \\
&= \sqrt{\frac{P_t}{N_t N_r}} \big(\mathbf{A}_1 \mathbf{H}_g s_1 + \mathbf{A}_2 \mathbf{H}_g s_2 + \cdots + \mathbf{A}_L \mathbf{H}_g s_L \big) + \mathbf{Z}.
\end{aligned}
\tag{A.1}
$$

Defining $\mathbf{h}_g = \mathrm{vec}(\mathbf{H}_g)$, $\mathbf{g} = \mathrm{vec}(\mathbf{G})$, and $\mathbf{h} = \mathrm{vec}(\mathbf{H})$, we define $\mathbf{b}_\ell$ as $\mathbf{b}_\ell = (\mathbf{I}_{N_r} \otimes \mathbf{A}_\ell) \mathbf{h}_g \triangleq (\mathbf{I}_{N_r} \otimes \mathbf{A}_\ell) (\mathbf{h} \circ \mathbf{g})$ for $\ell = 1, \ldots, L$. Then, vectorization of (A.1) results in

$$
\bar{\mathbf{y}} = \sqrt{\frac{P_t}{N_t N_r}} \big[\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_L \big] \mathbf{s} + \bar{\mathbf{z}},
\tag{A.2}
$$

where $\bar{\mathbf{y}} = \mathrm{vec}(\mathbf{Y})$ and $\bar{\mathbf{z}} = \mathrm{vec}(\mathbf{Z})$ are $TN_r \times 1$ vectors. Also, we define $\boldsymbol{\mathcal{H}}_g = \big(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_L \big)$ as the equivalent reconfigured channel matrix. $\boldsymbol{\mathcal{H}}_g$ is a real-valued matrix since $\mathbf{h} \circ \mathbf{g}$ represents the real vector $\mathbf{h}_g$ with entries equal to the amplitude of the channel coefficients. Matrix $\boldsymbol{\mathcal{H}}_g$ matrix exists when $\mathbf{H}_g$ is non-zero. Thus, (A.2) can be rewritten as $\bar{\mathbf{y}} = \sqrt{\frac{P_t}{N_t N_r}} \boldsymbol{\mathcal{H}}_g \mathbf{s} + \bar{\mathbf{z}}$. After applying the ZF receiver, the estimated $\hat{\mathbf{s}}_{\mathrm{ZF}}$ is given by

$$
\hat{\mathbf{s}}_{\mathrm{ZF}} = \sqrt{\frac{N_t N_r}{P_t}} \big(\boldsymbol{\mathcal{H}}_g^\dagger \boldsymbol{\mathcal{H}}_g \big)^{-1} \boldsymbol{\mathcal{H}}_g^\dagger \bar{\mathbf{y}} = \mathbf{s} + \sqrt{\frac{N_t N_r}{P_t}} \big(\boldsymbol{\mathcal{H}}_g^\dagger \boldsymbol{\mathcal{H}}_g \big)^{-1} \boldsymbol{\mathcal{H}}_g^\dagger \bar{\mathbf{z}}.
\tag{A.3}
$$

The existence of (A.3) depends on the existence of $\left(\boldsymbol{\mathcal{H}}_g^\dagger \boldsymbol{\mathcal{H}}_g\right)^{-1}$. Hence, $\boldsymbol{\mathcal{H}}_g^\dagger \boldsymbol{\mathcal{H}}_g$ should be full-rank. It is shown in in [133] that for any real-valued $n \times 1$ vector $\mathbf{v}$, we have

$$\mathbf{v}^\dagger \mathbf{A}_k^\dagger \mathbf{A}_\ell \mathbf{v} = \delta_{kl} \mathbf{v}^\dagger \mathbf{v}, \; k, \ell = 1, \ldots, L, \tag{A.4}$$

where $\delta_{k,l}$ is the Kronecker delta function and is 1 when $k = l$ and 0 when $k \neq l$. Considering that $\boldsymbol{\mathcal{H}}_g$ is real-valued and referring to (A.4), we can derive that

$$\boldsymbol{\mathcal{S}} \triangleq \boldsymbol{\mathcal{H}}_g^\dagger \boldsymbol{\mathcal{H}}_g \triangleq \|\mathbf{h}_g\|^2 \mathbf{I}_L. \tag{A.5}$$

Consequently, we have $\det(\boldsymbol{\mathcal{S}}) = \|\mathbf{h}_g\|^{2L}$, which is always non-zero. Eq. (A.5) indicates that the equivalent reconfigured channel matrix is orthogonal, which means the linear receivers achieve full-diversity gain [89] for the STBC defined in (3.9) under ideal phase-shifters.

A.2 Proof of Lemma 2

In the case of digital phase-shifters, the equivalent reconfigured channel matrix is expressed as $\tilde{\boldsymbol{\mathcal{H}}}_g = \left[\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_L\right]$, where $\tilde{\mathbf{b}}_\ell$ is given by

$$\tilde{\mathbf{b}}_\ell = \left(\mathbf{I}_{N_r} \otimes \mathbf{A}_\ell\right) \tilde{\mathbf{h}}_g \triangleq \left(\mathbf{I}_{N_r} \otimes \mathbf{A}_\ell\right) \left(\tilde{\mathbf{g}} \circ \mathbf{h}\right), \tag{A.6}$$

for $\ell = 1, \ldots, L$, with $\tilde{\mathbf{g}} = \text{vec}(\tilde{\mathbf{G}})$. The matrix $\tilde{\mathbf{G}}$ represents the receive lens antenna parameter for digital phase-shifters. The elements of $\tilde{\mathbf{G}}$ are defined as $\tilde{g}_{n_t, n_r} =$

$e^{-j\tilde{\theta}_{n_t,n_r}}$. The quantized phase $\tilde{\theta}_{n_t,n_r}$ is obtained as

$$\tilde{\theta}_{n_t,n_r} = \arg \min_{\theta_{n_t,n_r}} |\angle h_{n_t,n_r} - \theta_{n_t,n_r}|, \quad \text{subject to } \theta_{n_t,n_r} \in \{0, \ldots, \frac{(2^B-1)2\pi}{2^B}\}. \quad \text{(A.7)}$$

Since digital phase-shifters are not able to provide the exact value of the channel's phase, the vector $\tilde{\mathbf{b}}_\ell$ is not real-valued. It is simple to show that $\tilde{\mathbf{b}}_\ell^* \tilde{\mathbf{b}}_k \neq 0$, for $\ell \neq k$. Thus, the matrix $\tilde{\boldsymbol{\mathcal{H}}}_g^* \tilde{\boldsymbol{\mathcal{H}}}_g$ is not orthogonal. This would cast doubt on the existence of the ZF matrix in (A.3) when digital phase-shifters are employed. In order to study (A.3), we obtain $\boldsymbol{\mathcal{S}}$ in the case of digital phase shifting, $\tilde{\boldsymbol{\mathcal{S}}}$, as

$$\tilde{\boldsymbol{\mathcal{S}}} \triangleq \tilde{\boldsymbol{\mathcal{H}}}_g^* \tilde{\boldsymbol{\mathcal{H}}}_g = \begin{bmatrix} \|\mathbf{h}_g\|^2 & \tilde{\mathbf{b}}_1^* \tilde{\mathbf{b}}_2 & \cdots & \tilde{\mathbf{b}}_1^* \tilde{\mathbf{b}}_L \\ \tilde{\mathbf{b}}_2^* \tilde{\mathbf{b}}_1 & \|\mathbf{h}_g\|^2 & \cdots & \tilde{\mathbf{b}}_2^* \tilde{\mathbf{b}}_L \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{b}}_L^* \tilde{\mathbf{b}}_1 & \tilde{\mathbf{b}}_L^* \tilde{\mathbf{b}}_2 & \cdots & \|\mathbf{h}_g\|^2 \end{bmatrix}. \quad \text{(A.8)}$$

Each off-diagonal element in $\tilde{\boldsymbol{\mathcal{S}}}$ is given by

$$\tilde{\mathbf{b}}_k^* \tilde{\mathbf{b}}_\ell = \tilde{\mathbf{h}}_g^* \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g = \left(\tilde{\mathbf{h}}_g^{\text{Re}\dagger} - j \tilde{\mathbf{h}}_g^{\text{Im}\dagger} \right) \mathbf{A}_k^\dagger \mathbf{A}_\ell \left(\tilde{\mathbf{h}}_g^{\text{Re}} + j \tilde{\mathbf{h}}_g^{\text{Im}} \right)$$

$$\stackrel{(a)}{=} j \tilde{\mathbf{h}}_g^{\text{Re}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Im}} - j \tilde{\mathbf{h}}_g^{\text{Im}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Re}} \stackrel{(b)}{=} 2j \tilde{\mathbf{h}}_g^{\text{Re}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Im}}. \quad \text{(A.9)}$$

To obtain (a), for $\ell \neq k$, we notice that based on (A.4) we have $\tilde{\mathbf{h}}_g^{\text{Re}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Re}} = 0$ and $\tilde{\mathbf{h}}_g^{\text{Im}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Im}} = 0$. For Step (b), we rewrite the second term of (a) as $j \tilde{\mathbf{h}}_g^{\text{Im}\dagger} \mathbf{A}_\ell^\dagger \mathbf{A}_k \tilde{\mathbf{h}}_g^{\text{Re}}$. Noting that $\tilde{\mathbf{h}}_g^{\text{Im}\dagger} \mathbf{A}_\ell^\dagger \mathbf{A}_k \tilde{\mathbf{h}}_g^{\text{Re}}$ is a scalar real value, after applying the transpose operation, it can be rewritten as $\tilde{\mathbf{h}}_g^{\text{Re}\dagger} \mathbf{A}_k^\dagger \mathbf{A}_\ell \tilde{\mathbf{h}}_g^{\text{Im}}$. Thus, the two terms in (a) are the same and (b) is valid.

To continue from (A.9), we consider the property of the Hurwitz-Radon family of matrices where each column and row only contains one non-zero element which is either 1 or -1. in [86], it is revealed that $\mathbf{A}_k^\dagger \mathbf{A}_\ell$ is also the Hurwitz-Radon set of matrices, which means $\mathbf{A}_k^\dagger \mathbf{A}_\ell$ is skew-symmetric. Let (m,n) for $m,n = 1,2,\ldots,L$ denote the position of the non-zero element of the mth row and the nth column in matrix $\mathbf{A}_k^\dagger \mathbf{A}_\ell$. Since the matrix is skew-symmetric, for any non-zero element at (m,n), the element (n,m) is non-zero and its value is negative of the value of the element (m,n). Further, let vector $\mathbf{i}_{k,\ell}^1$ contain the first elements of all (m,n) pairs and vector $\mathbf{i}_{k,\ell}^2$ contain the second elements of all (m,n) pairs. For instance, let $\mathbf{A}_k^\dagger \mathbf{A}_\ell = [0\ 1\ 0\ 0; -1\ 0\ 0\ 0; 0\ 0\ 0\ -1; 0\ 0\ 1\ 0]$. In this 4×4 matrix, the non-zero elements are positioned in $(2,1)$, $(1,2)$, $(4,3)$, and $(3,4)$ which yields $\mathbf{i}_{k,\ell}^1 = (2,1,4,3)$ and $\mathbf{i}_{k,\ell}^2 = (1,2,3,4)$. Hence, it is not hard to see that for $k \neq \ell$

$$
\begin{aligned}
\tilde{\mathbf{b}}_k^* \tilde{\mathbf{b}}_\ell &= 2j \sum_{q=1}^{L} (-1)^{q+1} h_{g,\mathbf{i}_{k,\ell}^1(q)}^{\mathrm{Re}} h_{g,\mathbf{i}_{k,\ell}^2(q)}^{\mathrm{Im}} \\
&\overset{(c)}{=} 2j \sum_{q=1}^{L/2} \left(h_{g,\mathbf{i}_{k,\ell}^1(2q-1)}^{\mathrm{Re}} h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}^{\mathrm{Im}} - h_{g,\mathbf{i}_{k,\ell}^1(2q)}^{\mathrm{Re}} h_{g,\mathbf{i}_{k,\ell}^2(2q)}^{\mathrm{Im}} \right) \\
&\overset{(d)}{=} -2j \sum_{q=1}^{L/2} h_{g,\mathbf{i}_{k,\ell}^1(2q-1)} h_{g,\mathbf{i}_{k,\ell}^2(2q-1)} \sin\left(\Delta\delta_{k,\ell}\right).
\end{aligned}
\tag{A.10}
$$

Step (c) follows from the skew-symmetric property of $\mathbf{A}_k^\dagger \mathbf{A}_\ell$ which is shown in (3.10). To get (d), we replace $h_{g,\mathbf{i}_{k,\ell}^1(2q-1)}^{\mathrm{Re}}$ and $h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}^{\mathrm{Im}}$ by $h_{g,\mathbf{i}_{k,\ell}(2q-1)}\cos(\delta_{\mathbf{i}_{k,\ell}^1(2q-1)})$ and $h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}\sin(\delta_{\mathbf{i}_{k,\ell}^2(2q-1)})$, respectively, in which δ denotes the difference between the phase of the channel coefficient and the phase-shifter, then notice that $\sin x \cos y - \cos x \sin y = \sin(x-y)$. Also, we define $\Delta\delta_{k,\ell} = \delta_{\mathbf{i}_{k,\ell}^1(2q-1)} - \delta_{\mathbf{i}_{k,\ell}^2(2q-1)}$. The maximum absolute value of $\tilde{\mathbf{b}}_k^* \tilde{\mathbf{b}}_\ell$ is obtained when $\Delta\delta_{k,\ell} = 2\pi/2^{B+1}$. To take the resolution of the phase-shifters into account, we define a new matrix $\hat{\mathcal{S}}$ where the off-diagonal ele-

ments are denoted by $\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell$. For $\Delta\delta_{k,\ell} = \frac{2\pi}{2^{B+1}}$, the absolute value of each off-diagonal element is defined as

$$\left|\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell\right| = 2\sin\left(\frac{2\pi}{2^{B+1}}\right) \sum_{q=1}^{L/2} h_{g,\mathbf{i}_{k,\ell}^1(2q-1)} h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}. \tag{A.11}$$

We notice that each off-diagonal element is either $\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell = j \left|\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell\right|$ or $\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell = -j \left|\hat{\mathbf{b}}_k^* \hat{\mathbf{b}}_\ell\right|$. The diagonal elements are equal to $\tilde{\mathbf{b}}_\ell^* \tilde{\mathbf{b}}_\ell \triangleq \hat{\mathbf{b}}_\ell^* \hat{\mathbf{b}}_\ell = \|\mathbf{h}_g\|^2$. In what follows, we aim to find an upper bound for the off-diagonal elements of $\hat{\boldsymbol{\mathcal{S}}}$ which will be justified below. For a fixed phase-shifter resolution, the entries of $\hat{\boldsymbol{\mathcal{S}}}$ vary only with respect to the amplitude of the channel coefficients. Hence, we can normalize the amplitude of channel and rewrite $\hat{\boldsymbol{\mathcal{S}}}$ as $\hat{\boldsymbol{\mathcal{S}}} \triangleq \|\mathbf{h}_g\|^2 \bar{\bar{\boldsymbol{\mathcal{S}}}}$ in which the diagonal elements of $\bar{\bar{\boldsymbol{\mathcal{S}}}}$ are equal to 1 and absolute value of the upper bound of off-diagonal elements are given by

$$\left|\bar{\bar{\mathbf{b}}}_k^* \bar{\bar{\mathbf{b}}}_\ell\right| = \sin\left(\frac{2\pi}{2^{B+1}}\right) \frac{2\sum_{q=1}^{L/2} h_{g,\mathbf{i}_{k,\ell}^1(2q-1)} h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}}{\|\mathbf{h}_g\|^2} \leq \sin\left(\frac{2\pi}{2^{B+1}}\right). \tag{A.12}$$

The inequality is due to $\frac{2\sum_{q=1}^{L/2} h_{g,\mathbf{i}_{k,\ell}^1(2q-1)} h_{g,\mathbf{i}_{k,\ell}^2(2q-1)}}{\|\mathbf{h}_g\|^2} \leq 1$. To justify this, without loss of generality, we notice that for two arbitrary real scalars x and y the inequality $x^2 + y^2 \geq 2xy$ always holds.

The matrix $\bar{\bar{\boldsymbol{\mathcal{S}}}}$ is positive definite if $\text{Re}\left(\boldsymbol{v}^* \bar{\bar{\boldsymbol{\mathcal{S}}}} \boldsymbol{v}\right) > 0$ holds for all non-zero vectors

$\boldsymbol{v} \in \mathbb{C}^{L \times 1}$ in [134]. Hence, we have

$$
\begin{aligned}
\boldsymbol{v}^{*}\bar{\bar{\boldsymbol{\mathcal{S}}}}\boldsymbol{v} &= \sum_{i=1}^{L}|v_{i}|^{2} + \sum_{k=2}^{L}\sum_{\ell=1}^{k-1}\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\left(v_{\ell}v_{k}^{*} - v_{\ell}^{*}v_{k}\right) \\
&\overset{(a)}{=} \sum_{i=1}^{L}|v_{i}|^{2} - 2\sum_{k=2}^{L}\sum_{\ell=1}^{k-1}\left(\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right)^{\mathrm{Im}}\left(v_{\ell}^{\mathrm{Im}}v_{k}^{\mathrm{Re}} - v_{\ell}^{\mathrm{Re}}v_{k}^{\mathrm{Im}}\right) \\
&\overset{(b)}{=} \sum_{i=1}^{L}|v_{i}|^{2} - 2\sum_{k=2}^{L}\sum_{\ell=1}^{k-1}\left(\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right)^{\mathrm{Im}}|v_{k}||v_{\ell}|\sin(\theta_{\ell} - \theta_{k}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{L}|v_{i}|^{2} - 2\sum_{k=2}^{L}\sum_{\ell=1}^{k-1}\left|\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right||v_{k}||v_{\ell}| \\
&\overset{(d)}{\geq} \sum_{i=1}^{L}|v_{i}|^{2} - 2\sin\left(\frac{2\pi}{2^{B+1}}\right)\sum_{k=2}^{L}\sum_{\ell=1}^{k-1}|v_{k}||v_{\ell}| \overset{\triangle}{=} \boldsymbol{v}_{\mathrm{abs}}^{\dagger}\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}\boldsymbol{v}_{\mathrm{abs}}, \quad\quad (A.13)
\end{aligned}
$$

where $\boldsymbol{v}_{\mathrm{abs}}$ denotes the element-wise absolute value of $\boldsymbol{v}$ and $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is obtained by replacing the off-diagonal elements of $\bar{\bar{\boldsymbol{\mathcal{S}}}}$ by the upper bound in (A.12). From (a), since the terms $\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}$ and $v_{\ell}v_{k}^{*} - v_{\ell}^{*}v_{k}$ are purely imaginary, we get that $\boldsymbol{v}^{*}\bar{\bar{\boldsymbol{\mathcal{S}}}}\boldsymbol{v}$ is always real. For Step (b), we use $v_{\ell}^{\mathrm{Re}} = |v_{\ell}|\cos\theta_{\ell}$ and $v_{\ell}^{\mathrm{Im}} = |v_{\ell}|\sin\theta_{\ell}$ and then $\sin\theta_{\ell}\cos\theta_{k} - \cos\theta_{\ell}\sin\theta_{k} = \sin(\theta_{\ell} - \theta_{k})$. To obtain (c), we use the inequality $\left(\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right)^{\mathrm{Im}}\sin(\theta_{\ell} - \theta_{k}) \leq \left|\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right|$. Then, replacing $\left|\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right|$ by its maximum value $\left|\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right|$ and using the upper bound $\left|\bar{\bar{\mathbf{b}}}_{k}^{*}\bar{\bar{\mathbf{b}}}_{\ell}\right| \leq \sin\left(\frac{2\pi}{2^{B+1}}\right)$ defined in (A.12) give (d).

Ineq. (A.13) reveals that instead of the complicated matrix $\bar{\bar{\boldsymbol{\mathcal{S}}}}$, we can evaluate the positive definite property of $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$, i.e., if $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is positive definite, $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is positive definite as well. Note that the opposite is not true: $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ may not be full-rank while $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is full-rank. To evaluate the positive definiteness of $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ we use Sylvester's criterion in [135]. According to the criterion, if every leading principle minor of $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is positive, then $\bar{\bar{\boldsymbol{\mathcal{S}}}}_{\mathrm{upper}}$ is positive definite. The leading principle minor is defined as

the determinant of the leading principle sub-matrix. The leading principle sub-matrix of order p of an $L \times L$ matrix is obtained by deleting the last $L-p$ rows and columns of the matrix. Let $\bar{\bar{\mathbf{S}}}_{\text{upper},p}$ denote the leading principle sub-matrix of order p of $\bar{\bar{\mathbf{S}}}_{\text{upper}}$. For $p = 1$, it gives $\bar{\bar{\mathbf{S}}}_{\text{upper},1} = 1$. For $p > 1$, we have $\bar{\bar{\mathbf{S}}}_{\text{upper},p} = \left(1 - \sin(\frac{2\pi}{2^{B+1}})\right) \mathbf{I}_p + \boldsymbol{\omega}\boldsymbol{\omega}^\dagger$ with $\boldsymbol{\omega} = \sqrt{\sin(\frac{2\pi}{2^{B+1}})}\mathbf{1}$ in which the vector $\mathbf{1}$ is of size $p \times 1$ for $p = 2,\ldots,L$ where every element is equal to one. Then, we apply Sherman-Morrison formula in [136] to obtain the leading principle minor of $\bar{\bar{\mathbf{S}}}_{\text{upper},p}$ which gives

$$\det\left(\bar{\bar{\mathbf{S}}}_{\text{upper},p}\right) = \left(1 + (p-1)\sin\left(\frac{2\pi}{2^{B+1}}\right)\right)\left(1 - \sin\left(\frac{2\pi}{2^{B+1}}\right)\right)^{p-1}. \tag{A.14}$$

Having the leading principle minors in hand, one can determine that for what symbol size L and phase-shifter resolution B the matrix $\bar{\bar{\mathbf{S}}}_{\text{upper}}$ (respectively, $\bar{\bar{\mathbf{S}}}$) is positive definite, and hence the equivalent reconfigured channel matrix is full-rank. We note that for any symbol size and phase-shifter resolution $\det\left(\bar{\bar{\mathbf{S}}}_{\text{upper},1}\right) = 1 > 0$. Also, for any symbol size L and phase-shifter resolution $B = 1$, (A.14) becomes 0 for $p = 2,\ldots,L$. That is, $\bar{\bar{\mathbf{S}}}_{\text{upper}}$ is not positive definite. Consequently, the equivalent reconfigured channel matrix is not full-rank. Further, for any symbol size L and phase-shifter resolution $B > 1$, (A.14) is always positive for $p = 2,\ldots,L$ which means all the leading principle minors are positive. Therefore, $\bar{\bar{\mathbf{S}}}_{\text{upper}}$ is positive definite and the equivalent reconfigured channel matrix is full-rank. Although $\bar{\bar{\mathbf{S}}}_{\text{upper}}$ is positive definite, for some cases, its determinant is close to zero which may result in an ill-conditioned system. For instance, when $L = 4$ and $B = 2$, $\det\left(\bar{\bar{\mathbf{S}}}_{\text{upper}}\right) = 0.0784$ ($\det\left(\bar{\bar{\mathbf{S}}}_{\text{upper}}\right) = \det\left(\bar{\bar{\mathbf{S}}}_{\text{upper},L}\right)$). To resolve this issue, the resolution B should be chosen properly.

A.3   Proof of Theorem 2

Let assume that phase-shifter resolution B is chosen that the Hermitian matrix $\tilde{\boldsymbol{S}}$ defined in (A.8), and accordingly the normalized $\tilde{\boldsymbol{S}}$, is positive definite and its inverse exists as well. Here, to derive the upper bound, we start by finding an upper bound for $(\bar{\tilde{\boldsymbol{S}}}^{-1})_{\ell,\ell}$ as following. Let $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the maximum and the minimum eigenvalues of a matrix $(\cdot)$, respectively. Since $\bar{\tilde{\boldsymbol{S}}}^{-1}$ is positive definite, i.e., $\det(\bar{\tilde{\boldsymbol{S}}}^{-1}) \geq C$, we have $0 < \lambda_{\min}(\bar{\tilde{\boldsymbol{S}}}^{-1}) \leq (\bar{\tilde{\boldsymbol{S}}}^{-1})_{\ell,\ell} \leq \lambda_{\max}(\bar{\tilde{\boldsymbol{S}}}^{-1})$. So, we have

$$
\begin{aligned}
(\bar{\tilde{\boldsymbol{S}}}^{-1})_{\ell,\ell} \leq \lambda_{\max}(\bar{\tilde{\boldsymbol{S}}}^{-1}) &\overset{(a)}{=} \lambda_{\min}^{-1}(\bar{\tilde{\boldsymbol{S}}}) \\
&\overset{(b)}{\leq} \frac{1}{C}\left(\frac{\sum_{\ell=2}^{L}\lambda_\ell(\bar{\tilde{\boldsymbol{S}}})}{L-1}\right)^{L-1} \overset{(c)}{<} \frac{1}{C}\left(\frac{L}{L-1}\right)^{L-1}.
\end{aligned}
\tag{A.15}
$$

To get (a), since $\bar{\tilde{\boldsymbol{S}}}$ is a positive definite matrix, $\lambda_{\max}(\bar{\tilde{\boldsymbol{S}}}^{-1}) = \lambda_{\min}^{-1}(\bar{\tilde{\boldsymbol{S}}})$ holds in [89]. Step (b) is explained as follows. Noting $\prod_{\ell=1}^{L}\lambda_\ell(\bar{\tilde{\boldsymbol{S}}}) = \det(\bar{\tilde{\boldsymbol{S}}}) \geq C$, it gives $C\lambda_{\min}^{-1}(\bar{\tilde{\boldsymbol{S}}}) \leq \prod_{\ell=2}^{L}\lambda_\ell(\bar{\tilde{\boldsymbol{S}}})$. Applying the arithmetic-geometric inequality to $\prod_{\ell=2}^{L}\lambda_\ell(\bar{\tilde{\boldsymbol{S}}})$ results in (b). To get (c), we note that $\sum_{\ell=2}^{L}\lambda_\ell(\tilde{\boldsymbol{S}}) < \sum_{\ell=1}^{L}\lambda_\ell(\tilde{\boldsymbol{S}}) = \mathrm{tr}(\tilde{\boldsymbol{S}}) = L$.

Now, the received SNR of the ℓth symbol is given by

$$
\rho_\ell = \frac{P_t}{N_t N_r}\frac{1}{(\tilde{\boldsymbol{S}}^{-1})_{\ell,\ell}\sigma^2} > \kappa'\|\mathbf{h}_g\|^2,
\tag{A.16}
$$

in which $\kappa' = \frac{1}{C}a\rho\left(\frac{L}{L-1}\right)^{L-1}$, and a and ρ are defined in (3.11). Then, the SER expression is obtained as

$$
\mathrm{P}(\hat{s}_\ell \to s_\ell|\mathbf{h}_g) < \frac{M-1}{M}e^{-\kappa'\|\mathbf{h}_g\|^2}.
\tag{A.17}
$$

Following the same procedure in Theorem 1 in [89], it yields

$$\mathrm{P}(\hat{s}_\ell \to s_\ell) = \mathbb{E}_{\mathbf{h}_g}[\mathrm{P}(\hat{s}_\ell \to s_\ell | \mathbf{h}_g)] < \frac{M-1}{M} \kappa'^{-N_t N_r}.$$

A.4 Proof of Corollary 1

Recall that at each lens antenna the number of oriented beams is equal to the number of its selected TSAs. Since an independent AWGN noise is taken into account for each TSA, the noise term in (3.5), z_{t,n_r}, is comprised of the summation of all noise terms of those TSAs in each lens antenna, i.e., $z_{t,n_r} = \sum_{n_t=1}^{N_t} n_{t,n_r,n_t}$, where $n_{t,n_r,n_t} \sim \mathcal{CN}(0, \sigma_n^2)$. Due to the $i.i.d$ property of n_{t,n_r,n_t}, the variance of z_{t,n_r} is equal to $\sigma^2 = N_t \sigma_n^2$. Regarding (3.12), the SNR for the symbol s_ℓ is given by $\rho_\ell = \frac{P_t}{N_t N_r} \frac{\|\mathbf{h}_g\|^2}{\sigma^2} = \frac{P_t}{N_t N_r} \frac{\|\mathbf{h}_g\|^2}{N_t \sigma_n^2}$. Taking the average with respect to the channel vector, it yields $\rho = \frac{P_t}{N_t N_r} \frac{N_t N_r}{N_t \sigma_n^2} = \frac{P_t}{N_t \sigma_n^2}$.

A.5 Proof of Theorem 3

Let us start from (3.12) in Theorem 1. The outage probability is defined as

$$p_{\mathrm{out}} = 1 - \mathrm{P}(\rho_\ell > \gamma) = 1 - \mathrm{P}\left(\|\mathbf{h}_g\|^2 > \frac{\gamma \sigma^2 N_t N_r}{P_t}\right). \tag{A.18}$$

In practice, finding the exact PDF of $\|\mathbf{h}_g\|^2$ is not trivial. To the best of the authors' knowledge, two different types of techniques are available to approximate the PDF of the sum of variables with log-normal distribution. The first technique assumes that the PDF of the sum of log-normal variables is log-normal in [137, 138]. To systematically estimate the parameters of the log-normal distribution, i.e., μ' and η', four methods are proposed: Wilkinson in [137], Schwartz-Yeh in [137], Farley in [137],

and Cumulants matching in [138]. The main advantage of these methods is to systematically estimate the parameters. Although for the outage probability less than 0.1 these methods result in high estimation error in [139]. Other advantages and disadvantages of these methods are widely studied in [139]. The second technique relaxes the log-normal distribution assumption and focuses on finding accurate PDFs for a wide range of the outage probabilities. To this end, several methods such as minimax approximation in [140], least square approximation in [141], and log skew normal approximation in [142] are proposed. Due to the complicated PDF of these methods, calculating (A.18) is intractable. On the other hand, the log-normal distribution assumption makes the first technique attractive to use. Hence, we adopt a log-normal PDF for the variable $\|\mathbf{h}_g\|^2$. Since the systematic methods do not provide an accurate estimation for μ' and η', we do exhaustive search to find the best value of μ' and η'. With the log-normal assumption, the PDF of $10\log_{10}(\|\mathbf{h}_g\|^2)$ is normal with μ'-mean and η'^2-variance in dB. Hence, we have

$$p_{\text{out}} = 1 - \mathrm{P}\left(\|\mathbf{h}_g\|^2 > \frac{\gamma\sigma^2 N_t N_r}{P_t}\right) \overset{(a)}{=} 1 - \mathrm{P}\left(10\log_{10}(\|\mathbf{h}_g\|^2) > 10\log_{10}\left(\frac{\gamma\sigma^2 N_t N_r}{P_t}\right)\right)$$

$$(A.19)$$

To get (a), we apply $10\log_{10}$ to both sides and notice that the logarithm is an incremental function. Using the fact that $10\log_{10}(\|\mathbf{h}_g\|^2)$ closely has a normal distribution, (3.15) is obtained.

APPENDIX B:

PROOFS OF CHAPTER 3

B.1 Proof of Lemma 3

The minimum rate requirement constraint in (4.19c) imposes the following condition for p_i.

$$p_i \geq \underbrace{\left(\sum_{l=i+1}^{N_{\mathrm{RF}}} p_l + \frac{\sigma^2}{\alpha |h_{ik}|^2} \right)}_{(a)} \underbrace{\left(2^{\bar{R}_{ik}} - 1 \right)}_{(b)}. \tag{B.1}$$

We recall that users of the ith RAMA group are scheduled with almost equal channel gains, i.e, $|h_{i,k}| \approx |h_{i,k'}|$, for $k, k' \in \{1, \ldots, N_B\}$. Therefore, the term (a) has almost the same value for all users in one RAMA group. So, the minimum power of the ith RAMA group, ith RF chain, is lower bounded by the maximum of the term (b), i.e., $2^{\max_k \bar{R}_{ik}} - 1$.

APPENDIX C:

PROOFS OF CHAPTER 4

C.1   Proof of Theorem 4

Proof. Given the perfect alignment assumption and (5.11), the effective channel vector for $U_{n,m}$ becomes

$$\bar{\mathbf{h}}_{n,m}^{\dagger} = \beta_{n,m}\mathbf{a}_{\mathrm{BS}}^{\dagger}(\varphi_{n,m})\mathbf{F}_{\mathrm{RF}} = \beta_{n,m}\beta_{n,1}^{-1}\bar{\mathbf{h}}_{n,1}^{\dagger}. \tag{C.1}$$

On the other hand, we have

$$\bar{\mathbf{h}}_{n,1}^{\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell} = \begin{cases} \gamma_{n,n}, & \text{for } n, \ell = 1, 2, \ldots, N, \\ 0, & \text{for } \ell \neq n. \end{cases} \tag{C.2}$$

Therefore, using (C.1) and (C.2) the numerator in (5.4) becomes

$$P_{n,m}\left|\beta_{n,m}\right|^2\left|\beta_{n,1}\right|^{-2}\gamma_{n,n}^2. \tag{C.3}$$

Also, the intra-cluster interference in (5.5) becomes $I_{\mathrm{intra}}^{n,m} = \sum_{k=1}^{m-1} P_{n,k}\left|\beta_{n,m}\right|^2\left|\beta_{n,1}\right|^{-2}\gamma_{n,n}^2$, and the inter-cluster interference term becomes zero, i.e., $I_{\mathrm{inter}}^{n,m} = 0$.

Now, substituting (C.3), and the determined $I_{\text{intra}}^{n,m}$ and $I_{\text{inter}}^{n,m}$ in (5.4) gives

$$
\begin{aligned}
\overline{R}_{n,m} &= \log_2\left(1 + \frac{P_{n,m}\,|\beta_{n,m}|^2\,|\beta_{n,1}|^{-2}\,\gamma_{n,n}^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\,|\beta_{n,m}|^2\,|\beta_{n,1}|^{-2}\,\gamma_{n,n}^2 + \sigma^2}\right) \\[2mm]
&\stackrel{(a)}{=} \log_2\left(1 + \frac{P_{n,m}\,|\beta_{n,m}|^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\,|\beta_{n,m}|^2 + \sigma^2\left(\mathbf{F}^{-1}\right)_{n,n}}\right) \\[2mm]
&\stackrel{(b)}{\geq} \log_2\left(1 + \frac{P_{n,m}\,|\beta_{n,m}|^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\,|\beta_{n,m}|^2 + \sigma^2\kappa_{\min}^{-1}(\mathbf{F})}\right),
\end{aligned}
\tag{C.4}
$$

(a) follows by plugging (5.15) into the expression in the first line of (C.4) and using simple manipulations. To get (b), we note that $\mathbf{F}_{\text{RF}}$ is full-rank matrix which means $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{RF}}^\dagger$ is positive definite. Then, we have $\left(\mathbf{F}^{-1}\right)_{n,n} \leq \kappa_{\max}\left(\mathbf{F}^{-1}\right) = \kappa_{\min}^{-1}(\mathbf{F})$ in which $\kappa_{\max}(\cdot)$ and $\kappa_{\min}(\cdot)$ denote the maximum and minimum eigenvalues of $(\cdot)$. $\qquad\square$

C.2 Proof of Lemma 4

Proof. Suppose that the effective channel vectors are fed back by using infinite-resolution codebooks. Also, let $\hat{\mathbf{h}}_{n,m}$ denote the normalized effective channel vector for $\mathrm{U}_{n,m}$, i.e.,

$$
\hat{\mathbf{h}}_{n,m} = \frac{\tilde{\mathbf{h}}_{n,m}}{\|\tilde{\mathbf{h}}_{n,m}\|}.
\tag{C.5}
$$

The angle between two complex-valued vectors $\tilde{\mathbf{h}}_{n,m}$ and $\tilde{\mathbf{h}}_{n,1} \in V_{\mathbb{C}}$, denoted by Φ_{C}, is obtained as $\cos\Phi_{\mathrm{C}} \triangleq \rho_{n,m}e^{j\omega_{n,m}} = \hat{\mathbf{h}}_{n,1}^\dagger\hat{\mathbf{h}}_{n,m}$, where $(\rho_{n,m} \leq 1)$ is equal to

$\rho_{n,m} = \cos\Phi_{\mathrm{H}}(\hat{\tilde{\mathbf{h}}}_{n,1}, \hat{\tilde{\mathbf{h}}}_{n,m}) = \left|\hat{\tilde{\mathbf{h}}}_{n,1}^{\dagger}\hat{\tilde{\mathbf{h}}}_{n,m}\right|$, in which $\Phi_{\mathrm{H}}(\hat{\tilde{\mathbf{h}}}_{n,1}, \hat{\tilde{\mathbf{h}}}_{n,m})$, $0 \leq \Phi_{\mathrm{H}} \leq \frac{\pi}{2}$, is the Hermitian angle between two complex-valued vectors $\tilde{\mathbf{h}}_{n,1}$ and $\tilde{\mathbf{h}}_{n,m}$ and $\omega_{n,m}$, $-\pi \leq \omega_{n,m} \leq \pi$, is called their pseudo-angle [143]. The factor $\rho_{n,m}$ describes the angle between the two lines in the complex-valued vector space $V_{\mathbb{C}}$ [143].

To ease the analysis, the angle $\omega_{n,m}$ is neglected [143]. Hence, we find the angle between two lines which are defined by the two vectors $\hat{\tilde{\mathbf{h}}}_{n,1}$ and $\hat{\tilde{\mathbf{h}}}_{n,m}$. Considering these two vectors as two lines in the space $V_{\mathbb{C}}$ would be optimistic. However, the simulation results reveal that the derived misalignment model is still effective. Such that, the extracted lower bound for the sum-rate using the misalignment model is close to the exact value of the sum-rate.

For $\ell = n$, the misalignment factor $\rho_{n,m}$ can be calculated as

$$
\begin{aligned}
\rho_{n,m} &\triangleq \left|\hat{\tilde{\mathbf{h}}}_{n,1}^{\dagger}\hat{\tilde{\mathbf{h}}}_{n,m}\right| \stackrel{(a)}{=} \frac{\left|\beta_{n,m}\beta_{n,1}\mathbf{a}_{\mathrm{BS}}^{\dagger}(\varphi_{n,m})\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\dagger}\mathbf{a}_{\mathrm{BS}}(\varphi_{n,1})\right|}{\|\tilde{\mathbf{h}}_{n,m}\|\|\tilde{\mathbf{h}}_{n,1}\|} \\
&\stackrel{(b)}{=} \frac{\left|\beta_{n,m}\beta_{n,1}\mathbf{a}_{\mathrm{BS}}^{\dagger}(\varphi_{n,m})\mathbf{V}_1\boldsymbol{\Lambda}_1\mathbf{V}_1^{\dagger}\mathbf{a}_{\mathrm{BS}}(\varphi_{n,1})\right|}{\|\tilde{\mathbf{h}}_{n,m}\|\|\tilde{\mathbf{h}}_{n,1}\|} \\
&\stackrel{(c)}{=} \frac{\displaystyle\sum_{i=1}^{N}\kappa_i\left|\mathbf{a}_{\mathrm{BS}}^{\dagger}(\varphi_{n,m})\mathbf{v}_1^i\mathbf{v}_1^{i\dagger}\mathbf{a}_{\mathrm{BS}}(\varphi_{n,1})\right|}{\sqrt{\displaystyle\sum_{\ell=1}^{N}K_{N_{\mathrm{BS}}}(\varphi_{\ell,1}-\varphi_{n,m})}\sqrt{\displaystyle\sum_{\ell=1}^{N}K_{N_{\mathrm{BS}}}(\varphi_{\ell,1}-\varphi_{n,1})}}.
\end{aligned}
\tag{C.6}
$$

To get (a), the expression in (5.11) is used. To get (b), we apply SVD to the Hermitian matrix $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\dagger}$ which gives $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\dagger} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\dagger}$ where $\mathbf{V}$ of size $N_{\mathrm{BS}} \times N_{\mathrm{BS}}$ is a unitary matrix and $\boldsymbol{\Lambda}$ of size $N_{\mathrm{BS}} \times N_{\mathrm{BS}}$ is a diagonal matrix of singular values ordered in

decreasing order. We then partition two matrices $\mathbf{V}$ and $\mathbf{\Lambda}$ as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{C.7}$$

where $\mathbf{V}_1$ is of size $N_{\mathrm{BS}} \times N$ and $\mathbf{\Lambda}_1$ and is of size $N \times N$. We note that $\mathrm{rank}(\mathbf{F}_{\mathrm{RF}}) = N$. Term (c) follows from the fact that $\mathbf{\Lambda}_1$ is a diagonal matrix with elements κ_i for $i = 1, 2, \ldots, N$. Notice that $\mathbf{v}_1^i$ represents the ith column.

For $\ell \neq n$, it is reasonable to assume that $\sqrt{1 - \rho_{n,m}^2}$ percentage of the amplitude of $\tilde{\mathbf{h}}_{n,m}$ leakages into the subspace generated by the other first users. To determine the subspace, we start with considering the impact of the misalignment imposed by the other first users on $\mathrm{U}_{n,m}$, i.e, $\sum_{\ell=1, \ell \neq n}^{N} \left| \tilde{\mathbf{h}}_{\ell,1}^\dagger \tilde{\mathbf{h}}_{n,m} \right|^2$. Using the definition of vector norm, we rewrite this expression as following:

$$\sum_{\ell=1, \ell \neq n}^{N} \left| \tilde{\mathbf{h}}_{\ell,1}^\dagger \tilde{\mathbf{h}}_{n,m} \right|^2 = \left\| \tilde{\mathbf{h}}_{n,m}^\dagger \begin{bmatrix} \tilde{\mathbf{h}}_{1,1} & \cdots & \tilde{\mathbf{h}}_{n-1,1} & \tilde{\mathbf{h}}_{n+1,1} & \cdots & \tilde{\mathbf{h}}_{N,1} \end{bmatrix} \right\|^2$$

$$\overset{(a)}{=} \left\| \tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{F}_{\mathrm{RF}}^\dagger \begin{bmatrix} \beta_{1,1} \mathbf{a}_{\mathrm{BS}} \left(\varphi_{1,1} \right) & \cdots & \beta_{n-1,1} \mathbf{a}_{\mathrm{BS}} \left(\varphi_{n-1,1} \right) \right.$$

$$\left. \beta_{n+1,1} \mathbf{a}_{\mathrm{BS}} \left(\varphi_{n+1,1} \right) \quad \cdots \quad \beta_{N,1} \mathbf{a}_{\mathrm{BS}} \left(\varphi_{N,1} \right) \right] \right\|^2$$

$$\overset{(b)}{=} \left\| \tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{F}_{\mathrm{RF}}^\dagger \mathbf{A}_{\mathrm{BS}}^{-n} \right\|^2. \tag{C.8}$$

To get (a), we replace $\tilde{\mathbf{h}}_{\ell,1}$ by (5.11). Since $\mathbf{a}_{\mathrm{BS}} \left(\varphi_{n,1} \right)$s are independent vectors, $\mathbf{G}_{\mathrm{BS}}^{-n} = \mathbf{F}_{\mathrm{RF}}^\dagger \mathbf{A}_{\mathrm{BS}}^{-n}$ determines an $N - 1$ dimensional subspace. We represent the weighted linear combination of $\hat{\tilde{\mathbf{h}}}_{\ell,1}^\dagger$ by a new vector $\mathbf{g}_{\mathrm{BS}}^{-n}$ which is located in the subspace $\mathbf{G}_{\mathrm{BS}}^{-n}$. So, we get $\mathbf{g}_{\mathrm{BS}}^{-n} = \mathbf{F}_{\mathrm{RF}} \times \sum_{\ell=1, \ell \neq n}^{N} \sqrt{P_\ell} \beta_{\ell,1} \mathbf{a}_{\mathrm{BS}}(\varphi_{\ell,1})$. To get (5.23), we only need to normalize $\mathbf{g}_{\mathrm{BS}}^{-n}$. $\qquad \square$

C.3   Proof of Theorem 5

Proof. Using (5.23), we obtain the following expressions. First,

$$
\left|\tilde{\mathbf{h}}_{n,m}^{\dagger}\mathbf{f}_{\mathrm{BB}}^{n}\right|^{2} = \rho_{n,m}^{2}\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2}\left|\hat{\tilde{\mathbf{h}}}_{n,1}^{\dagger}\mathbf{f}_{\mathrm{BB}}^{n}\right|^{2} + \left(1-\rho_{n,m}^{2}\right)\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2}\left|\mathbf{g}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{n}\right|^{2}
$$

$$
\overset{(a)}{=} \rho_{n,m}^{2}\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2}\left|\hat{\tilde{\mathbf{h}}}_{n,1}^{\dagger}\mathbf{f}_{\mathrm{BB}}^{n}\right|^{2} \overset{(b)}{=} \rho_{n,m}^{2}\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2}\left\|\tilde{\mathbf{h}}_{n,1}\right\|^{-2}\gamma_{n,n}^{2}, \tag{C.9}
$$

in which (a) follows since $\mathbf{g}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{n} = 0$ and (b) follows from (C.2). Second,

$$
\left|\tilde{\mathbf{h}}_{n,m}^{\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell}\right|^{2} = \left(1-\rho_{n,m}^{2}\right)\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{2}\left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell}\right|^{2}, \quad \text{for} \quad \ell \neq n. \tag{C.10}
$$

Next, Using (5.15), (C.2), (5.22), (C.5), and (C.9), (5.5) becomes

$$
I_{\mathrm{intra}}^{n,m} = \sum_{k=1}^{m-1} P_{n,k}\rho_{n,m}^{2}\left|\beta_{n,m}\right|^{2}\left(\mathbf{F}^{-1}\right)_{n,n}^{-1}K_{N_{\mathrm{BS}},m}K_{N_{\mathrm{BS}},1}^{-1}, \tag{C.11}
$$

where $K_{N_{\mathrm{BS}},1}$ and $K_{N_{\mathrm{BS}},m}$ are defined in (5.26). Likewise,using (C.2), (5.22), (C.5), and (C.10), (5.6) becomes

$$
I_{\mathrm{inter}}^{n,m} = \left(1-\rho_{n,m}^{2}\right)\left|\beta_{n,m}\right|^{2}\sum_{\ell \neq n}^{N} P_{\ell}\left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell}\right|^{2}K_{N_{\mathrm{BS}},m}. \tag{C.12}
$$

Further, after substituting (C.9), (C.11) and (C.12) into (5.4), we get

$$
\tilde{R}_{n,m} = \log_{2}\left(1+\frac{\Psi}{I_{\mathrm{intra}}^{n,m}+I_{\mathrm{inter}}^{n,m}+\sigma^{2}}\right) \overset{(a)}{\geq} \log_{2}\left(1+\frac{\Psi}{I_{\mathrm{intra}}^{n,m}+\varsigma_{\mathrm{inter}}^{n,m}+\sigma^{2}}\right), \tag{C.13}
$$

where $\Psi = P_{n,m}\rho_{n,m}^{2}\left|\beta_{n,m}\right|^{2}\left(\mathbf{F}^{-1}\right)_{n,n}^{-1}K_{N_{\mathrm{BS}},m}K_{N_{\mathrm{BS}},1}^{-1}$, and
$\varsigma_{\mathrm{inter}}^{n,m} = \left(1-\rho_{n,m}^{2}\right)\left|\beta_{n,m}\right|^{2}\times\kappa_{\max}(\mathbf{S})K_{N_{\mathrm{BS}},m}$. To get (a), we have the following lemma.

Lemma 5. An upper bound of $\sum_{\ell=1,\ell\neq n}^{N} P_\ell \left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell}\right|^2$ is the maximum eigenvalue of $\mathbf{S}$, i.e., $\kappa_{\max}(\mathbf{S})$.

Proof. We rewrite $\sum_{\ell=1,\ell\neq n}^{N} P_\ell \left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{f}_{\mathrm{BB}}^{\ell}\right|^2 = \left\|\mathbf{g}_{\mathrm{BS}}^{-n\dagger}\mathbf{F}_{\mathrm{BB}}^{-n,W}\right\|_2^2$. Maximizing $\left\|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{F}_{\mathrm{BB}}^{-n,W}\right\|_2^2$ given $\left\|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n}\right\| = 1$ is similar to maximizing a beamforming vector for maximum ratio transmission systems [144, 145]. Hence, the maximum value of $\hat{\mathbf{g}}_{\mathrm{BS}}^{-n}$ is the dominant right singular vector of $\mathbf{F}_{\mathrm{BB}}^{-n,W}$ [144, 145]. Thus, the maximum of $\left\|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger}\mathbf{F}_{\mathrm{BB}}^{-n,W}\right\|_2^2$ is equal to the maximum eigenvalue of $\mathbf{S}$. $\qquad\square$

Lemma 5 indicates that $I_{\mathrm{inter}}^{n,m} \leq \varsigma_{\mathrm{inter}}^{n,m}$. After some manipulations

$$
\begin{aligned}
\tilde{R}_{n,m} &\geq \log_2\left(1 + \frac{P_{n,m}\rho_{n,m}^2 |\beta_{n,m}|^2}{\zeta_{\mathrm{intra}}^{n,m} + (\varsigma_{\mathrm{inter}}^{n,m} + \sigma^2)(\mathbf{F}^{-1})_{n,n} K_{N_{\mathrm{BS}},m}^{-1} K_{N_{\mathrm{BS}},1}}\right) \\
&\overset{(a)}{\geq} \log_2\left(1 + \frac{P_{n,m}\rho_{n,m}^2 |\beta_{n,m}|^2}{\zeta_{\mathrm{intra}}^{n,m} + \zeta_{\mathrm{inter}}^{n,m} + \sigma^2\kappa_{\min}^{-1}(\mathbf{F})K_{N_{\mathrm{BS}},m}^{-1} K_{N_{\mathrm{BS}},1}}\right),
\end{aligned} \tag{C.14}
$$

where in the first line, $\zeta_{\mathrm{intra}}^{n,m} = \sum_{k=1}^{m-1} P_{n,k}\rho_{n,m}^2 |\beta_{n,m}|^2$ and in the second line, $\zeta_{\mathrm{inter}}^{n,m} = \left(1 - \rho_{n,m}^2\right) \times N_{\mathrm{BS}}N_{\mathrm{U}} |\beta_{n,m}|^2 \kappa_{\max}(\mathbf{S})\kappa_{\min}^{-1}(\mathbf{F})K_{N_{\mathrm{BS}},1}$. To get (a), we note that $(\mathbf{F}^{-1})_{n,n} \leq \kappa_{\min}^{-1}(\mathbf{F})$. $\qquad\square$

C.4   Proof of Theorem 6

Proof. We start with (5.4) to define the achievable rate of $U_{n,m}$ for the perfect correlation and the imperfect correlation, i.e., $\overline{R}_{n,m}$ and $\tilde{R}_{n,m}$, respectively. This gives

$$
\begin{aligned}
\Delta R_{n,m} &\triangleq \overline{R}_{n,m} - \tilde{R}_{n,m} \\[2mm]
&= \log_2\left(1 + \frac{P_{n,m}\left|\overline{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\left|\overline{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sigma^2}\right) - \\[2mm]
&\qquad \log_2\left(1 + \frac{P_{n,m}\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sum_{\ell=1,\ell\neq n}^{N} P_\ell\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^\ell\right|^2 + \sigma^2}\right) \\[2mm]
&\overset{(a)}{\leq} \log_2\left(\frac{\displaystyle\sum_{k=1}^{m} P_{n,k}\left|\overline{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sigma^2}{\displaystyle\sum_{k=1}^{m} P_{n,k}\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sigma^2}\right) - \\[2mm]
&\qquad \log_2\left(\frac{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\left|\overline{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sigma^2}{\displaystyle\sum_{k=1}^{m-1} P_{n,k}\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + \sum_{\ell=1,\ell\neq n}^{N} P_\ell\left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^\ell\right|^2 + \sigma^2}\right) \\[2mm]
&\overset{(b)}{\leq} \log_2\left(\frac{\left\|\overline{\mathbf{h}}_{n,m}\right\|^2 \left|\hat{\overline{\mathbf{h}}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2}{\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2 \left|\hat{\tilde{\mathbf{h}}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2}\right) - \log_2\left(\frac{\left\|\overline{\mathbf{h}}_{n,m}\right\|^2 \displaystyle\sum_{k=1}^{m-1} P_{n,k}\left|\hat{\overline{\mathbf{h}}}_{n,m}^\dagger \mathbf{f}_{\text{BB}}^n\right|^2 + 1}{\Upsilon}\right),
\end{aligned}
$$

$$(C.15)$$

where $\Upsilon = \left\|\tilde{\mathbf{h}}_{n,m}\right\|^2 \sum_{k=1}^{m-1} P_{n,k} \left|\hat{\tilde{\mathbf{h}}}_{n,m}^\dagger \mathbf{f}_{\mathrm{BB}}^n\right|^2 + \left\|\tilde{\mathbf{h}}_{n,m}\right\|^2 \sum_{\ell=1,\ell\neq n}^N P_\ell \left|\hat{\tilde{\mathbf{h}}}_{n,m}^\dagger \mathbf{f}_{\mathrm{BB}}^\ell\right|^2 + \sigma^2$. To get

(a) we remove positive quantity $\sum_{\ell=1,\ell\neq n}^N P_\ell \left|\tilde{\mathbf{h}}_{n,m}^\dagger \mathbf{f}_{\mathrm{BB}}^\ell\right|^2$ from the second term. Then, we

exchange the denominator of the first term with the numerator of the second one. (b)

follows from the fact that for $u > v$, it gives $\log\left(\frac{u}{v}\right) > \log\left(\frac{u+c}{v+c}\right)$ $(c > 0)$, and applying

the normalized vector $\tilde{\mathbf{h}}_{n,m}$ defined in (C.5) for both perfect and imperfect effective

channel vectors.

Noting that $\hat{\bar{\mathbf{h}}}_{n,1} = \hat{\bar{\mathbf{h}}}_{n,m}$ and using (C.9) it yields

$$\Delta R \leq \log_2\left(\frac{\left\|\bar{\mathbf{h}}_{n,m}\right\|^2}{\rho_{n,m}^2\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2}\right) - \log_2\left(\sum_{k=1}^{m-1} P_{n,k}\left\|\bar{\mathbf{h}}_{n,m}\right\|^2 \left|\hat{\bar{\mathbf{h}}}_{n,1}^\dagger \mathbf{f}_{\mathrm{BB}}^n\right|^2 + \sigma^2\right)$$

$$+ \log_2\left(\sum_{k=1}^{m-1} P_{n,k}\rho_{n,m}^2\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2 \left|\hat{\bar{\mathbf{h}}}_{n,1}^\dagger \mathbf{f}_{\mathrm{BB}}^n\right|^2 + (1 - \rho_{n,m}^2)\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2 \sum_{\ell=1,\ell\neq n}^N P_\ell \left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger} \mathbf{f}_{\mathrm{BB}}^\ell\right|^2 + \sigma^2\right)$$

$$\overset{(a)}{=} -\log_2\left(\sum_{k=1}^{m-1} P_{n,k}\rho_{n,m}^2 \left|\hat{\bar{\mathbf{h}}}_{n,1}^\dagger \mathbf{f}_{\mathrm{BB}}^n\right|^2\right)$$

$$+ \log_2\left(\sum_{k=1}^{m-1} P_{n,k}\rho_{n,m}^2 \left|\hat{\bar{\mathbf{h}}}_{n,1}^\dagger \mathbf{f}_{\mathrm{BB}}^n\right|^2 + (1 - \rho_{n,m}^2) \sum_{\ell=1,\ell\neq n}^N P_\ell \left|\hat{\mathbf{g}}_{\mathrm{BS}}^{-n\dagger} \mathbf{f}_{\mathrm{BB}}^\ell\right|^2 + \frac{\sigma^2}{\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2}\right)$$

$$\overset{(b)}{\leq} \log_2\left(1 + \frac{\left(1 - \rho_{n,m}^2\right)\kappa_{\max}(\mathbf{S}) + \sigma^2\left\|\tilde{\mathbf{h}}_{n,m}\right\|^{-2}}{\rho_{n,m}^2 K_{N_{\mathrm{BS}},1}^{-1}\left(\mathbf{F}^{-1}\right)_{n,n}^{-1}\sum_{k=1}^{m-1} P_{n,k}}\right)$$

$$\overset{(c)}{\leq} \log_2\left(1 + \frac{\left(1 - \rho_{n,m}^2\right)\kappa_{\max}(\mathbf{S}) + \sigma^2 K_{N_{\mathrm{BS}},m}^{-1} N_{\mathrm{BS}}^{-1} N_{\mathrm{U}}^{-1}\left|\beta_{n,m}\right|^{-2}}{\rho_{n,m}^2 K_{N_{\mathrm{BS}},1}^{-1}\kappa_{\min}(\mathbf{F})\sum_{k=1}^{m-1} P_{n,k}}\right), \qquad \text{(C.16)}$$

in which (a) follows by rewriting the first term as $\log_2\left(\rho_{n,m}^{-2}\left\|\bar{\mathbf{h}}_{n,m}\right\|^2\right) - \log_2\left(\left\|\tilde{\mathbf{h}}_{n,m}\right\|^2\right)$.

Then, we sum up the expression $\log_2\left(\rho_{n,m}^{-2}\|\bar{\mathbf{h}}_{n,m}\|^2\right)$ with the second term and the expression $-\log_2\left(\|\tilde{\mathbf{h}}_{n,m}\|^2\right)$ with the third term. To get (b), we again sum up the first term with the second term. We then use Lemma 5 to get $\kappa_{\max}(\mathbf{S})$ and (C.2) and (5.15) to get $K_{N_{\mathrm{BS}},1}^{-1}\,(\mathbf{F}^{-1})_{n,n}^{-1}$. To obtain (c), first we use $\|\tilde{\mathbf{h}}_{n,m}\|^2 = K_{N_{\mathrm{BS}},m}|\beta_{n,m}|^2$. Next we use the inequality $(\mathbf{F}^{-1})_{n,n} \leq \kappa_{\min}^{-1}(\mathbf{F})$. $\qquad\square$

APPENDIX D:

PROOFS OF CHAPTER 5

D.1   Proof of Proposition 1

Since $F(g_1,f_1)$ is a continuous function defined on a bounded closed set, it has a maximum point according to the extreme value theorem. Also, according to the critical point theorem, the maximum point should either be a stationary point or a boundary point. It is easy to observe that $F(g_1,f_1)$ is a monotonic increasing function of g_1 and a monotonic decreasing function of f_1. This means $\frac{\partial F}{\partial g_1} > 0$ and $\frac{\partial F}{\partial f_1} < 0$, i.e., there is no stationary point for $F(g_1,f_1)$ on the defined domain. Then, the maximum point should lie on the five boundaries: (i) $f_1 = lb_f$, $g_1 \in [lb_g,lb_f^2]$, (ii) $g_1 = lb_g$, $f_1 \in [lb_f,ub_f]$, (iii) $f_1 = ub_f$, $g_1 \in [lb_g,ub_g]$, (iv) $g_1 = ub_g$, $f_1 \in [\sqrt{ub_g},ub_f]$, and (v) $g_1 = f_1^2$, $f_1 \in [lb_f,\sqrt{ub_g}]$.

For the boundary (i), since $F(g_1,f_1)$ is a monotonic increasing function of g_1, the maximum point can only lie on the point (lb_f,lb_f^2) which belongs to the boundary (v) as well. For the boundary (ii), since $F(g_1,f_1)$ is a monotonic increasing function of g_1, we can pick $g_1 > lb_g$ to increase the value of $F(g_1,f_1)$. This implies that the maximum point cannot lie on the boundary (ii). Similarly, the maximum point cannot lie on the boundary (iii) either. For the boundary (iv), since $F(g_1,f_1)$ is a monotonic decreasing function of f_1, the maximum point can only lie on the point $(\sqrt{ub_g},ub_g)$ which belongs to boundary (v) as well. Note that the possible maximum points on the boundaries (i) and (iv) also belong to the boundary (v). Therefore, the maximum point must lie on the boundary (v) and the proof is complete.

D.2   Proof of Proposition 2

Since $G(g_1,f_1)$ is a continuous function defined on a bounded closed set, it has a maximum point according to the extreme value theorem. Also, according to the

critical point theorem, the maximum point should either be a stationary point or a boundary point. It is easy to observe that $G(g_1, f_1)$ is a monotonic decreasing function of f_1. $\frac{\partial G}{\partial f_1} < 0$, i.e., there is no stationary point for $G(g_1, f_1)$ on the defined domain. Then, the maximum point should lie on the four boundaries: (i) $g_1 = lb_g$, $f_1 \in [\sqrt{lb_g}, ub_f]$, (ii) $f_1 = ub_f$, $g_1 \in [lb_g, ub_g]$, (iii) $g_1 = ub_g$, $f_1 \in [\sqrt{ub_g}, ub_f]$, and (iv) $g_1 = lb_f^2$, $f_1 \in [\sqrt{lb_g}, \sqrt{ub_g}]$.

For the boundary (i), since $G(g_1, f_1)$ is a monotonic decreasing function of f_1, the maximum point can only lie on the point $(\sqrt{lb_g}, lb_g)$ which belongs to the boundary (iv) as well. For the boundary (ii), since $G(g_1, f_1)$ is a monotonic decreasing function of f_1, we can pick $f_1 < ub_f$ to increase the value of $G(g_1, f_1)$, which implies that the maximum point cannot lie on this boundary. For the boundary (iii), since $G(g_1, f_1)$ is a monotonic decreasing function of f_1, the maximum point must lie on the point $(\sqrt{ub_g}, ub_g)$, which also belongs to the boundary (iv). Since the possible maximum points on the boundaries (i) and (iii) also belong to the boundary (iv), the maximum point must lie on the boundary (iv) and the proof is complete.