UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

ARTHUR SOUSA DE SENA

DESIGN AND PERFORMANCE ENHANCEMENT OF MULTI-CLUSTER MULTI-USER MASSIVE MIMO-NOMA NETWORKS

FORTALEZA

2019

ARTHUR SOUSA DE SENA

DESIGN AND PERFORMANCE ENHANCEMENT OF MULTI-CLUSTER MULTI-USER MASSIVE MIMO-NOMA NETWORKS

Dissertação apresentada à coordenação do Programa de Pós-Graduação em Engenharia de Teleinformática da Universidade Federal do Ceará, como requisito parcial para a obtenção do título de Mestre em Engenharia de Teleinformática. Área de concentração: Sinais e Sistemas.

Orientador: Prof. Dr. Daniel Benevides da Costa.

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S477d Sena, Arthur.
Design and Performance Enhancement of Multi-Cluster Multi-User Massive MIMO-NOMA Networks / Arthur Sena. – 2019.
98 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2019.
Orientação: Prof. Dr. Daniel Benevides da Costa.

1. Non-orthogonal multiple access (NOMA). 2. massive MIMO. 3. multi-polarized antennas. 4. successive sub-array activation. I. Título.

CDD 621.38

ARTHUR SOUSA DE SENA

DESIGN AND PERFORMANCE ENHANCEMENT OF MULTI-CLUSTER MULTI-USER MASSIVE MIMO-NOMA NETWORKS

Dissertation presented to the Coordination of the Teleinformatics Engineering Post-Graduation Program of the Universidade Federal do Ceará as a part of the requisites to obtain the Master's Degree in Teleinformatics Engineering. Concentration Area: Signals and Systems.

Approved in: 26/08/2019.

COMMITTEE

Prof. Dr. Daniel Benevides da Costa (Advisor)
Universidade Federal do Ceará (UFC)

Prof. Dr. Rafael Timóteo de Sousa Júnior
Universidade de Brasília (UnB)

Prof. Dr. Walter da Cruz Freitas Júnior
Universidade Federal do Ceará (UFC)

I dedicate this dissertation to my beloved father, who is now watching over me from heaven. I will always miss you.

ACKNOWLEDGMENTS

In the first place, I thank God, for always blessing my life and enlightening my path. I am very grateful for all the achievements that God has already provided me, and for all that is yet to come.

I would like to express my heartfelt gratitude to my beloved girlfriend, Ana Jéssica da Silva, for her incessant love, affection, patience and for always being by my side, even in the most difficult moments. Her love and support were crucial for all accomplishments achieved during my master's studies.

I would like to express my deepest gratitude to my parents, José Pereira de Sena and Mirian Pereira de Sousa, who have provided me endless support and encouragement during my entire life. Sadly, my father could not see the completion of this dissertation, but, wherever he is now, I am sure that he is very proud of his son achieving the Master's degree. Not less important, I would like to thank my aunt and second mother, Antonia Miraci de Sousa, for her constant optimism and solicitude. I also would like to thank my aunt Erotildes Pereira Paiva, for being such a kind and devoted person, and my dear sister, Aline Maria Sousa de Sena, for making my days more joyful. You all are essential in my life and responsible, in large part, for my achievements.

I would like to express my deepest thanks to my advisor, Prof. Dr. Daniel Benevides da Costa, for always helping me in everything I needed, and for all the opportunities he has offered me. His commitment, competence, and constant assistance have been essential for the development of this research. I also would like to express my gratitude to Prof. Dr. Zhiguo Ding, and Prof. Dr. Pedro Henrique Juliano Nardelli, for the important contributions and feedbacks.

My sincere thanks to the committee members, Prof. Dr. Rafael Timóteo de Sousa Júnior, and Prof. Dr. Walter da Cruz Freitas Júnior, for the availability, the time dedicated to the evaluation of this work and for the valuable suggestions.

At last, but not least, I would like to express my thanks to all the friends who have contributed in some way to the completion of this degree. Especially, to my friends Brena Kelly, Jair Alves, João Rafael, Leonardo Paiva, and Syllas Rangel. You have made this journey more pleasant and fun.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“God provides the wind, but man must raise the sails.” Augustine of Hippo

RESUMO

NOMA (*non-orthogonal multiple access*) e MIMO (*multiple-input multiple-output*) massivo são duas tecnologias essenciais para atender os exigentes requisitos da quinta geração (5G) das redes de comunicação sem fio. Especificamente, através de um número massivo de antenas, o MIMO massivo explora o domínio espacial para atender múltiplos usuários em paralelo. A técnica NOMA também serve múltiplos usuários simultaneamente, mas diferentemente do MIMO, a transmissão paralela é realizada através do domínio da potência, em que os usuários utilizam SIC (*successive interference cancellation*) para efetuar a recepção. A junção de MIMO massivo e NOMA consegue fornecer ganhos de desempenho ainda maiores. No entanto, devido a vários problemas, como pequeno espaçamento entre antenas, alta correlação espacial e desvanecimento profundo em canais sem fio, o desempenho do sistema ainda pode ser degradado e a confiabilidade da comunicação afetada, o que não é desejado nas exigentes redes 5G. Felizmente, foi demonstrado que a instalação de antenas com polaridades ortogonais pode aliviar as limitações de espaço entre as antenas e os problemas de correlação. O uso de estratégias de diversidade também é uma maneira eficaz de mitigar os efeitos prejudiciais dos canais de desvanecimento. Neste contexto, esta dissertação visa melhorar o desempenho de redes MIMO-NOMA massivas através da aplicação de multi-polarização e de técnicas de diversidade. Na primeira parte, projeta-se e avalia-se o desempenho de sistemas *multi-cluster* MIMO-NOMA massivo equipados com antenas de dupla polarização. Considera-se que uma única estação base (BS, do inglês *base station*) se comunica com múltiplos usuários em modo *downlink* e que todos os terminais também são equipados com múltiplas antenas de dupla polarização. Em particular, dois pré-codificadores são propostos: (i) o primeiro deles maximiza o número de grupos de usuários que são atendidos simultaneamente dentro de um *cluster*; e (ii) o segundo fornece maiores ganhos de desempenho em relação ao primeiro através da exploração de diversidade de polarização. Na segunda parte desta dissertação, propõe-se um novo esquema de diversidade para os sistemas MIMO-NOMA massivo, chamado de ativação sucessiva de *sub-arrays* (SSAA, do inglês *successive sub-array activation*). Para o sistema operando com SSAA, considerando um cenário com múltiplos *clusters*, em que a BS envia símbolos redundantes através de múltiplos *sub-arrays* para usuários com múltiplas antenas, propõe-se um pré-codificador de baixa complexidade construído apenas com as informações estatísticas do canal. Para todos os sistemas propostos (MIMO-NOMA com dupla polarização e SSAA), realizam-se estudos analíticos aprofundados, nos quais derivam-se expressões em forma fechada para as probabilidades de *outage*, e obtém-se as respectivas aproximações para altos valores de razão sinal-ruído (SNR, do inglês *signal-to-noise ratio*). Baseado nas análises assintóticas, as ordens de diversidade dos sistemas são determinadas. Expressões para as somas de taxas ergódicas também são derivadas. Resultados numéricos e de simulação são fornecidos para validar a análise analítica e demonstrar a superioridade de desempenho dos sistemas propostos.

Palavras-chave: NOMA, MIMO massivo, antennas multi-polarizadas, ativação sucessiva de *sub-arrays*.

ABSTRACT

Non-orthogonal multiple access (NOMA) and massive multiple-input multiple-output (MIMO) have arisen as essential enabling technologies for meeting the demanding requisites of the fifth-generation (5G) of wireless communication networks. Specifically, massive MIMO explores the space domain through a massive number of antennas to serve multiple users in parallel, while NOMA can also serve multiple users simultaneously, but differently from MIMO, the parallel transmission is performed by multiplexing the users in the power domain, in which successive interference cancellation (SIC) is employed for reception. The combination of massive MIMO and NOMA can provide even further performance improvements. However, due to various impairments, such as closely spaced antennas, high spatial correlation, or deep fading in wireless channels, the system performance can still be degraded and the communication reliability impacted, which is not desired in the demanding 5G networks. Fortunately, it has been demonstrated that the installation of co-located orthogonal polarized antennas can alleviate the inter-antenna space limitations and correlation issues. The use of diversity strategies is also an effective way of mitigating the harmful effects of fading channels. In this context, this dissertation aims to enhance the performance of multi-cluster multi-user massive MIMO-NOMA networks through the application of both multi-polarization and diversity techniques. First, we design and evaluate the performance of dual-polarized massive MIMO-NOMA systems, in which a single base station communicates in downlink mode with multiple users, with all terminals being equipped with multiple co-located dual-polarized antennas. In particular, two precoder designs are proposed: (i) the first one aims to maximize the number of user groups that are simultaneously served within a cluster; and (ii) the second approach aims to provide further improvements compared to the first one by exploring polarization diversity. In the second part of this dissertation, we propose a novel successive sub-array activation (SSAA) diversity scheme for a single-polarized massive MIMO-NOMA system, in which we also consider the downlink mode. In this scenario, the base station sends redundant symbols through multiple transmit sub-arrays to multi-antenna receivers, based on which a low-complexity two-stage precoder, that is constructed based only on the long-term channel statistical information, is proposed. For both dual-polarized and SSAA systems, we carry out in-depth analytical studies, in which closed-form expressions for the outage probabilities are derived. High signal-to-noise ratio (SNR) outage approximations are obtained, and the systems diversity orders are determined. The ergodic sum-rates are also derived. Numerical and simulation results are provided to validate the analytical analysis and to demonstrate the performance superiority of the proposed designs.

Keywords: Non-orthogonal multiple access (NOMA), massive MIMO, multi-polarized antennas, successive sub-array activation.

LIST OF FIGURES

ACRONYMS

3GPP	3rd Generation Partnership Project
5G	Fifth-generation
BPCU	Bits per Channel Use
BS	Base Station
CDF	Cumulative Distribution Function
CoMP	Coordinated Multi-Point
i.i.d.	independent and identically distributed
LOS	Line-of-Sight
LTE-A	Long-Term Evolution Advanced
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
NLOS	Non-Line-of-Sight
NOMA	Non-Orthogonal Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PDF	Probability Density Function
Q-PSK	Quadrature Phase-Shift Keying
RF	Radio Frequency
SC	Superposition Coding
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SSAA	Successive Sub-Array Activation
STBC	Space-Time Block Coding
SVD	Singular Value Decomposition
TDMA	Time Division Multiple Access
XPD	Cross-Polar Discrimination
XPI	Cross-Polar Isolation
XPR	Cross-Polar Ratio

LIST OF SYMBOLS

$\mathbb{R}$	Field of real numbers
$\mathbb{C}$	Field of complex numbers
a	Scalar
$\mathbf{a}$	Column vector
$\mathbf{A}$	Matrix
$\|\mathbf{a}\|$	Norm of $\mathbf{a}$
$[\mathbf{a}]_i$	i-th entry of $\mathbf{a}$
$[\mathbf{A}]_{i,j}$	(i,j) entry of $\mathbf{A}$
$[\mathbf{A}]_{i,*}$	i-th row of $\mathbf{A}$
$\mathbf{A}^T$	Transpose of $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of $\mathbf{A}$
$\mathbf{A}^\dagger$	Moore-Penrose pseudo-inverse of $\mathbf{A}$
$\mathrm{tr}\{\mathbf{A}\}$	Trace of $\mathbf{A}$
$\mathbf{I_M}$	Identity matrix of dimension $M \times M$
$\mathbf{0}_{M\times N}$	Matrix of dimension $M \times N$ with all zero entries
$\mathbf{1}_{M\times N}$	Matrix of dimension $M \times N$ with all one entries
$\odot$	Hadamard product
$\otimes$	Kronecker product
$\mathbb{E}[\cdot]$	Expectation
$\lfloor\cdot\rfloor$	Floor function
$\Gamma(\cdot)$	Gamma function
$\gamma(\cdot,\cdot)$	Lower incomplete Gamma function
$\mathrm{Ei}(\cdot)$	Exponential integral
$\mathbf{A} \sim \mathscr{W}_M^{-1}(N, \mathbf{B}^{-1})$	Inverse Wishart distributed matrix of dimension $M \times M$ with N degrees of freedom and covariance matrix $\mathbf{B}^{-1}$

CONTENTS

1 INTRODUCTION

1.1 Context and Motivations

Advances in communication technology are leading the human race to a hyper-connected society, where everyone and everything is going to interact with each other. This major evolution will trigger an unprecedented explosion in the number of connected devices, which is expected to reach more than 100 billion connections by 2030 [1]. In this new era, unforeseen services and applications will arise, from smart-homes and self-driving cars to interconnected low-power sensors in agriculture and factories. These new deployments will impose the most diverse requirements to the wireless systems, such as massive connectivity, low latency, and reliable communications [2, 3, 4]. In order to attend these highly demanding requisites, numerous strategies and technologies have been proposed to be incorporated in the fifth-generation (5G) of wireless communications. In particular, non-orthogonal multiple access (NOMA) has arisen as an essential enabling technique which has the capability of increasing the spectral efficiency and, at the same time, to reduce the system latency. These attractive features are accomplished by exploring the power domain to multiplex different users, while sharing the same resource blocks [5, 6, 7]. Briefly speaking, a base station (BS) superposes the messages of each user through superposition coding (SC) by assigning distinct power allocation coefficients and, at the receiver side, each user recovers its message by employing successive interference cancellation (SIC) technique. Due to the potential improvements that it achieves, NOMA was included in the 3rd generation partnership project (3GPP) long-term evolution advanced (LTE-A) Release-13 under the name of multi-user superposed transmission (MUST), in the 3GPP Release-14, where fifteen NOMA uplink schemes were proposed [5] and, more recently, in the specification for 5G New Radio standard, introduced in Release 15 [8, 9].

Massive multiple-input multiple-output (MIMO) is another enabling technology for 5G that boosts the capacity of communication systems without requiring additional bandwidth. Specifically, massive MIMO explores the space domain through a very large number of antennas to communicate with multiple users in parallel. The performance gains achieved in MIMO systems are scaled with the increase in the antenna array size, such that, as the number of transmit antennas goes to infinity, the channel matrices of different users become asymptotically orthogonal [10, 11, 12]. This condition, commonly called of favorable propagation, makes linear processing, such as zero-forcing precoding, to achieve near-optimum performance. Also, with the employment of large-scale arrays, due to a controlled superposition of wavefronts achieved by precoding, the radiated energy can be concentrated within extremely narrow areas around the receiver. This feature can remarkably increase the energy efficiency of the system and improve the coverage range of transmitted signals [13]. As a consequence of this efficient signal propagation, costs with hardware components can also be reduced. For instance, the expensive 50W amplifiers commonly used in conventional systems can be replaced by multiple low-cost

low-power amplifiers [13]. Due to these great advantages, the massive MIMO technology has been also included in 3GPP releases 13, 14 and 15 [8, 9].

In combination with NOMA, massive MIMO can provide even higher spectral and connectivity improvements that can remarkably outperform conventional schemes employing orthogonal multiple access (OMA) protocols [6, 7]. However, a reliable communication cannot always be guaranteed. Due to various reasons, the system performance can be degraded and the communication quality impacted. Especially in 5G networks, due to the use of higher frequency bands (above 6 GHz), the penetration loss and atmospheric absorption can strongly affect the system reliability [14]. High co-channel interference in ultra-dense networks, which is expected to be common in 5G deployments, is also very problematic and has a strong impact on the system performance [15]. The high mobility scenario is another critical application, where the fast-varying channels and double-selective fading make it very difficult to establish a good communication [16]. In real NOMA deployments, the SIC reception, if not well designed, can lead to error propagation and impact the stability of decoding [17]. Also, employing a massive number of antennas can introduce some problems. For instance, the strong channel correlation caused by the closely spaced antennas can severely deteriorate the performance of MIMO systems [18, 19, 20].

The above limitations have been extensively investigated in classical MIMO networks for decades. For example, it has been demonstrated that the issue related to the inter-antenna space in massive antenna arrays can be efficiently alleviated by installing co-located orthogonal polarized antennas [18, 19, 20, 21, 22]. The referred strategy enables the design of compact massive arrays with a very low correlation between orthogonal antenna elements. For the other mentioned hostile scenarios, an effective solution for overcoming performance degradation can be achieved by employing diversity techniques, which has been shown to dramatically improve the performance of communication systems under the inhospitality of fading channels [23]. Nevertheless, when it comes to massive MIMO-NOMA systems, the number of contributions available in the literature, related to both multi-polarization and diversity, is extremely limited. This lack of relevant contributions makes evident that there is still a lot to investigate and develop in order to deploy efficient and reliable massive MIMO-NOMA systems capable of fulfilling the stringent requirements of 5G. Given the aforementioned motivations, this dissertation aims to study and propose strategies for enhancing the performance of massive MIMO-NOMA networks. The main contributions of this work are organized into two chapters. In Chapter 3, for the first time in the literature, we design and perform an in-depth analysis of massive MIMO-NOMA systems with dual-polarized antenna arrays, and, in Chapter 4, a novel successive sub-array activation (SSAA) diversity scheme is proposed to enhance the performance of each user in a single-polarized MIMO-NOMA network. Specific details about our contributions are given in the referred chapters. Next, a review of the most relevant related works is presented.

1.2 Related Works

1.2.1 Single-Polarized Antennas in MIMO-NOMA Systems

Considering a single cell deployment, Ding *et al.* [6] employed NOMA to a downlink multi-user MIMO scenario assuming fewer antennas at the BS than at the users. An in-depth analytical analysis was carried out and closed-form outage expressions were derived. In [7], a multi-user multi-cluster MIMO-NOMA system was investigated assuming that the number of transmit antennas was higher than the ones at the users. In this work, the authors proposed a limited feedback scheme for the users' ordering information and provided an analytical outage analysis. The multi-cell MIMO-NOMA case was considered in [24], where path-following optimization algorithms were proposed to maximize the overall sum throughput. The designed precoders provided better average sum throughput than conventional schemes. Cooperative MIMO-NOMA networks have been addressed in [25, 26, 27]. A new non-regenerative massive MIMO-NOMA relay system design was proposed in [25], in which a closed-form expression for the signal-to-interference-plus-noise ratio (SINR) was derived. This novel relay model outperformed conventional NOMA and OMA cooperative systems. In [26], the authors proposed a suboptimal algorithm to maximize the cell-edge users' achievable rate, and in [27] the performance of a multi-relay massive MIMO-NOMA system was analyzed. The spectrum and energy efficiencies of a NOMA heterogeneous network were studied in [28].

1.2.2 Multi-Polarized Antennas in MIMO-NOMA Systems

The design of polarized schemes has gained increasing attention along the last years. In [18], Park *et al.* proposed dual-stage precoders based on both spatial correlation and antenna polarization for a single-antenna multi-user massive MIMO scenario. The authors provided an asymptotic performance analysis for the proposed precoding strategies and showed that the dual-polarized system outperforms the single-polarized in terms of sum-rate. However, only the BS was considered to be equipped with dual-polarized antennas. In [21, 22], a 3D geometrical model was proposed for dual-polarized MIMO systems, in which it was shown that polarized MIMO exhibits higher capacity performance and robustness when considering Ricean channels. Although the authors validated the proposed model through data field measurements, an analytical analysis was not provided. Multi-polarization has also been considered in numerous codebook design works for LTE-A MIMO systems [19, 29, 30, 31]. The employment of polarized antennas in MIMO-NOMA was only considered in [32]. However, in the refereed paper, a 3D triple-polarized beamforming scheme was proposed to reduce inter-beam interference, which is completely different from the goals of this work. Besides, only simulation results were presented, lacking analytical derivation.

1.2.3 Diversity in MIMO-NOMA systems

A few works have addressed diversity in massive MIMO-NOMA setups. In [33], antenna diversity was employed to enhance the outage performance of a downlink MIMO-NOMA system with users equipped with only two receive antennas, and the expansion to a scenario with multiple receive antennas was proposed in [34]. In both works, the authors provided closed-form expressions for the outage probability and also analyzed the system ergodic rates. However, in the proposed designs, only one NOMA group is served at a time, which is a limitation. Considering a single base station with only two transmit antennas, the work in [35] achieves diversity through the combination of Alamouti space-time block coding (STBC) and MIMO-NOMA, in which a high signal-to-noise ratio (SNR) approximation and an exact expression for the outage probability was derived. In [36], a multi-dimensional STBC was proposed to a MIMO sparse code multiple access system, for two and four transmit antennas. The application of STBC to cooperative NOMA networks was investigated in [37] and [38]. Coordinated multi-point (CoMP) is employed to MIMO-NOMA systems in [39] and [40] to improve the bit error rate and energy efficiency of cell-edge users. Performance improvements in MIMO-NOMA systems can also be achieved through antenna selection schemes [41, 42]. However, it was shown that the optimal solution is very complex and, since NOMA sends a unique superposed symbol, the antenna selection can not maximize the SINR of all users at the same time.

1.3 Publications

- **Conference Papers**
  1. **SENA, A. S.**; COSTA, D. B.; DING, Z.; NARDELLI, P. H. J. On the Performance of Massive MIMO-NOMA Networks with Dual-Polarized Antenna Array. In: **IEEE Global Communications Conference (GLOBECOM)**, Waikoloa, USA, Dec. 2019. **(Accepted)**
  2. **SENA, A. S.**; COSTA, D. B.; DING, Z.; NARDELLI, P. H. J.; DIAS, U. S. Successive Sub-Array Activation for Massive MIMO-NOMA Networks. In: **IEEE International Conference on Communications (ICC)**, Dublin, Ireland, Jun. 2020. (To be submitted)
- **Journal Papers**
  1. **SENA, A. S.**; COSTA, D. B.; DING, Z.; NARDELLI, P. H. J. Massive MIMO-NOMA Networks with Multi-Polarized Antennas. **IEEE Transactions on Wireless Communications**. **(Accepted)**
  2. **SENA, A. S.**; COSTA, D. B.; DING, Z.; NARDELLI, P. H. J.; DIAS, U. S.; PAPADIAS, C. B. Massive MIMO-NOMA Networks with Successive Sub-Array Activation. **IEEE Transactions on Wireless Communications**. (Under review)

1.4 Dissertation Structure

The remainder of this dissertation is organized as follows.

- **Chapter 2**
  In this chapter, basic concepts and important theoretical background necessary for Chapters 3 and 4 are introduced. We present the main features of NOMA, provide a detailed explanation of massive MIMO technology, discuss the combination of NOMA and massive MIMO, describe the spatial correlation model adopted in this work, and introduce details of relevant precoding and detection techniques. Concepts of multi-polarization and diversity strategies are also presented.

- **Chapter 3**
  The first part of our main contributions are presented in this chapter, in which we design and analyze the performance of a multi-cluster multi-user dual-polarized massive MIMO-NOMA systems. In-depth analytical studies are provided, which are followed by insightful numerical and simulation results. For instance, our results show that the proposed dual-polarized MIMO-NOMA designs outperform conventional single-polarized systems, even for high cross-polar interference.

- **Chapter 4**
  This chapter brings the second part of our contributions, in which a novel diversity strategy, named successive sub-array activation (SSAA), is proposed for massive MIMO-NOMA systems. A full theoretical analysis and representative numerical and simulation results are provided. Our results show that the proposed system operating with SSAA outperforms conventional full array massive MIMO setups.

- **Chapter 5**
  Lastly, this chapter summarizes the main contributions accomplished in this dissertation, provides the final considerations, and discusses some promising directions for future works.

2 FUNDAMENTAL BACKGROUND

In this chapter, fundamental concepts and relevant theoretical background related to the contributions given in this dissertation are introduced. First, in Section 2.1, we highlight the main advantages of NOMA and present the principles of SC and SIC. Then, in Section 2.2, a comprehensive explanation of massive MIMO is presented, in which some attractive features and challenges of the technology are discussed. The spatial correlation model employed in this work is also introduced. Besides, we provide concepts of precoding techniques and show details for the application of NOMA in massive MIMO, which are followed by the explanation of the detection strategy adopted in this study. In Section 2.3, important concepts of multi-polarization are introduced, and Section 2.4 concludes the chapter with the basics of diversity.

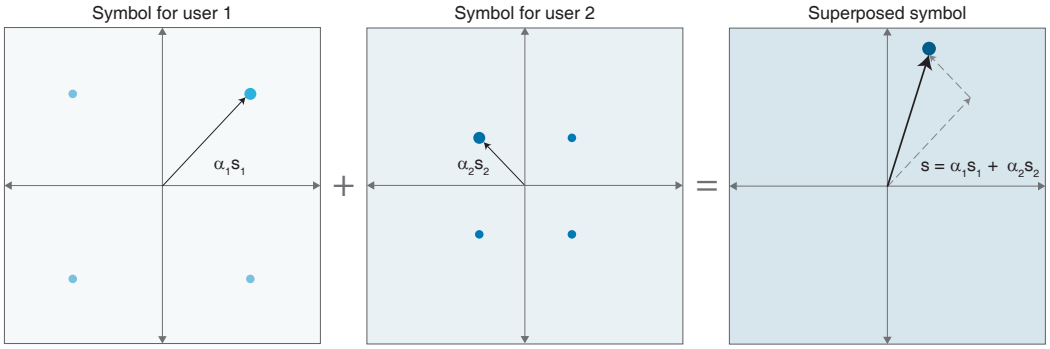
2.1 Non-Orthogonal Multiple Access

As briefly discussed in Section 1.1, power-domain NOMA enables multiple users to be served in parallel within the same resource block. The additional degrees of freedom introduced by the technique scales the spectral efficiency of communication systems up to a new patamar, which is essential for supporting the massive number of connections expected in the upcoming years. This makes NOMA one of the most accredited and promising techniques for enabling 5G requisites. In addition to what has been mentioned, NOMA can offer many other attractive advantages, such as compatibility with the current generation technology, improved interference management, low latency, and enhanced user fairness. More details about these benefits are summarized next.

2.1.1 Main Benefits of NOMA

- **Compatibility –** Since an independent new dimension is explored, i.e., the power domain, NOMA is considered to be an add-on technique, which is compatible with the current technology employed in commercial communication systems [43, 44, 45]. Therefore, for the practical implementation of NOMA, no drastic changes in legacy network architecture are required. In addition to that, it has been shown that NOMA can work in cooperation with conventional OMA techniques such as time division multiple access (TDMA) and orthogonal frequency division multiple access (OFDMA) [43]. For example, by employing NOMA in OFDMA, it is possible to serve multiple users on a same subcarrier.
- **Improved interference management –** Although OMA techniques can perfectly eliminate intra-cell interference, they are not very suitable for the massive connectivity of 5G. For instance, since in OMA the interference management is conventionally concentrated entirely at the BS, a large number of connections would lead to a huge processing overload. This issue is efficiently addressed by NOMA, in which part of the interfer-

Figure 1: Superposition coding process. Two Q-PSK symbols are superposed.



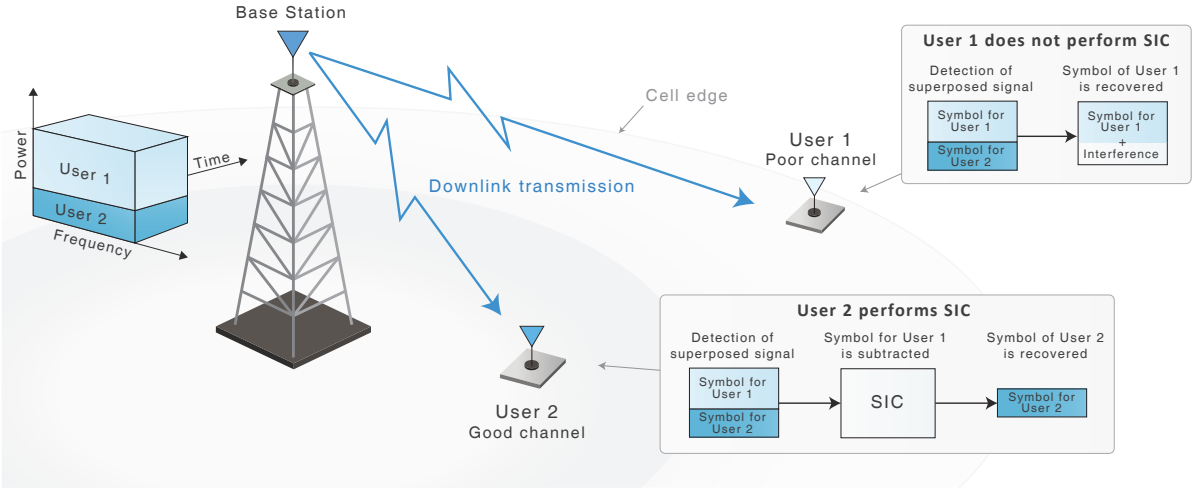
Source: Created by the author.

ence cancellation is performed in the users' devices by employing SIC. It is also worth mentioning that, even though SIC slightly increases the computational complexity of reception at the users' devices, strategies such as user grouping can be used to reduce the complexity of signal decoding [46].

- **Low latency –** In conventional OMA, the number of orthogonal spectrum resources is very limited, and the OMA users are only able to transmit when an empty resource block is available. This makes the system latency to become considerable high, which gets even worse when the number of connections rises [47]. Therefore, OMA can not satisfy the stringent latency requisites of 5G. On the other hand, NOMA is capable of serving simultaneously several users within the same resource block. Thus, NOMA users do not need to wait for a scheduled time slot or subcarrier in order to receive or transmit information. This beneficial feature results in a remarkable reduction in transmission latency [48].
- **Enhanced user fairness –** In OMA, independently of the channel conditions, an entire resource block is granted to each user. However, users with poor channel quality demand much less spectrum than the stronger ones. As a consequence, in situations where fairness is required, in which weak users needs to be served, the scarce spectrum resources will be inefficiently occupied. The use of OMA in these scenarios will impact the overall performance of the system and reduce the sum-rate [43]. Then, OMA is definitely not efficient when the objective is to provide fairness. In contrast, as more details will be provided in the next subsection, NOMA can serve in parallel both weak and strong users while sharing the same resource. This capability is very beneficial to the system performance, in which the high rates of strong users maintain the system sum-rate high at the same time that low-rate users are served for fairness.

In order to further understand how users are multiplexed in power-domain NOMA, next, we present a detailed explanation about SC and SIC technologies.

Figure 2: Illustration of a two-user downlink NOMA network.



Source: Created by the author.

2.1.2 Superposition Coding

As previously stated, NOMA relies on two key technologies, namely SC and SIC. In particular, SC can be viewed as a multiple layer modulation technique, in which a super-symbol comprises multiple sub-symbols intended for different users [49]. The basic concept of SC employed in NOMA, consists of superposing multiple symbols by assigning different power levels, in which, conventionally, more power is allocated to users with worse channel conditions and less power to the ones with good channel conditions. With this approach, even users near the cell-edge are able to achieve satisfactory throughput [44], which naturally improves fairness within the network. Figure 1 illustrates the SC process for the simple case when two quadrature phase-shift keying (Q-PSK) symbols intended for two different users are superposed. As one can observe, following the SC principle, more power has been allocated to the user 1 (the weaker user). For the two-user scenario, it has been demonstrated that SC is capable of achieving the capacity region of degraded Gaussian channels [50, 51, 52]. Furthermore, when the number of symbols becomes sufficiently large, i.e., when the number of symbols approaches infinity, SC ideally approaches the Shannon capacity [49].

2.1.3 Successive Interference Cancellation

Considering the downlink mode, while SC is performed at the BS, SIC is carried out at the users' devices. In simple terms, SIC consists of a multi-user detection technique that, based on the received power, sequentially detects the users' symbols. Specifically, during the SIC process, the users first decode the symbol with the highest power (the symbol intended for the weakest user) while treating the lower-power symbols as interference. Then, the decoded data achieved during the first decoding phase is subtracted from the received superposed signal. The same procedure is performed to the symbol with the second highest power, and for the succeeding symbols as well. This iterative decoding process repeats until the current user decodes

its own symbol.

To better illustrate the decoding process of SIC, let us consider the simple two-user scenario, as shown in Figure 2. User 1 represents a cell-edge user with poor channel condition, and User 2 denotes a cell-center user with strong channel condition. Then, assume that users 1 and 2 receive, respectively, the following superposed signals: $y_1 = h_1(\alpha_1 s_1 + \alpha_2 s_2)$ and $y_2 = h_2(\alpha_1 s_1 + \alpha_2 s_2)$, in which h_i repents the channel gain, α_i is the power allocation coefficient, and s_i is the data symbol, $i = 1,2$. Moreover, since $|h_1|^2 < |h_2|^2$, following the SC principle, the power allocated at the BS satisfies $|\alpha_1|^2 > |\alpha_2|^2$. For this simple case, considering a noiseless signal model, the decoding process can be summarized as follows.

- **Signal detection at User 1:**
  1. Decode s_1 from y_1 by treating s_2 as interference.
- **Signal detection at User 2:**
  1. Decode s_1 from y_2 by treating s_2 as interference.
  2. Subtract from y_2 the decoded signal of previous step, e.g. $y_2 - h_2\alpha_1 s_1$.
  3. Decode s_2 without any interference.

As can be realized, in the two-user scenario, the weak user will suffer from interference, while the strong one will recover its symbol without any corruption. This behavior also happens in the scenario with more than two users, in which the weakest user receives interference from everyone. The best user, ideally, is not affected by interference, but it needs to perform a high number of SIC decodings, which can become excessively complex as the number of users increases. Due to this characteristic, SIC is considered to be an interference-limited technique. Consequently, the number of users served in NOMA should be maintained low in order to achieve a good system performance [53]. Despite this impairment, SIC can provide some advantages. For instance, in [54], it was stated that SIC can achieve considerable better performance and lower hardware complexity than other interference cancellation techniques. In addition, it has been demonstrated in [55, 23] that SIC can approach the Shannon capacity in both single-antenna and MIMO systems.

2.2 Massive Multiple-Input Multiple-Output

The idea of equipping both BS and receivers with multiple antennas to improve the capacity of wireless channels dates back to the 1990s [56, 57, 58]. However, only very recently, in 2010 [59], a large number of BS antennas was considered in a multi-user scenario. The refereed work showed for the first time the potential spectral improvements that large-scale antenna arrays can provide, giving birth to what is known today as massive MIMO technology. Since then, massive MIMO gained tremendous popularity among both the research community and industry, and it is confirmed to be one of the main technologies in 5G [9]. Since 2016, numerous telecommunication companies around the world have been performing trials with massive MIMO, and commercial transceivers are already available [60].

In addition to the great spectral improvements that can be achieved, there are two worth mentioning phenomena that happen when the number of BS antennas goes to infinity, namely channel hardening and favorable propagation [59]. In order to explain these properties, let us consider a practical scenario. Suppose that a single BS equipped with M antennas is communicating in downlink mode with multiple single-antenna users, in which $\mathbf{h}_u$ represents the channel vector of user u. Then, the aforementioned phenomena can be explained as follows:

- **Channel hardening –** The effects of small-scale fading become deterministic. More specifically, for a large number of transmit antennas ($M \to \infty$), the channel gain $\|\mathbf{h}_u\|^2$ converges to its expected value, i.e., $\|\mathbf{h}_u\|^2 \to \mathbb{E}[\|\mathbf{h}_u\|^2]$.
- **Favorable propagation –** The propagation channels of different users become orthogonal for $M \to \infty$, meaning that $\mathbf{h}_u^H \mathbf{h}_{u'}$=0, for $u' \neq u$.

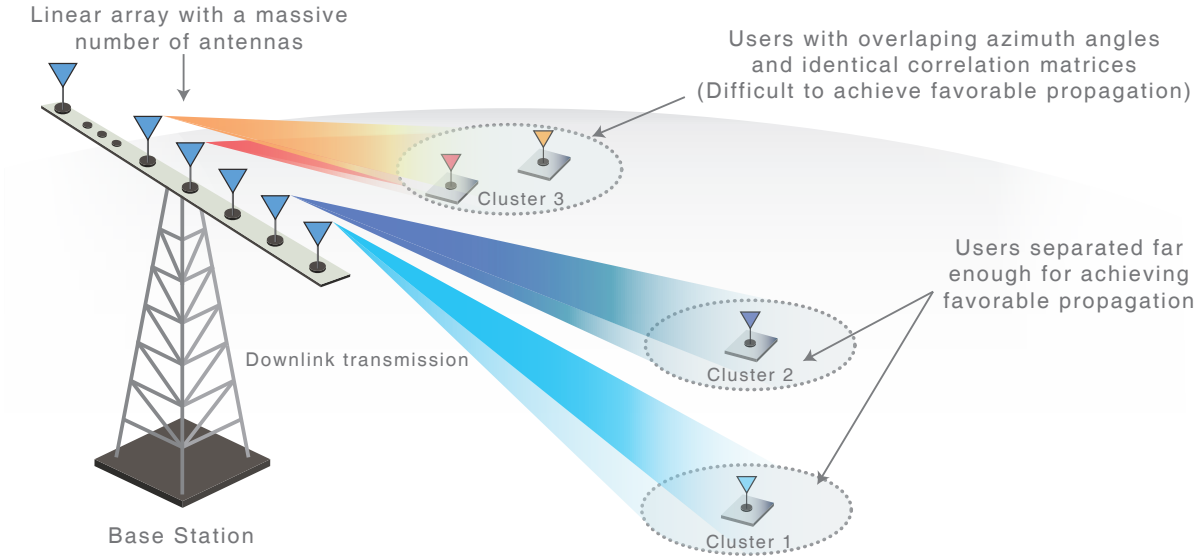
These properties show that the employment of a large number of antennas makes easier to mitigate interference among multiple users. In addition, the adoption of simple linear precoders becomes enough to reach satisfactory performance levels. These results have been used and confirmed in various succeeding works [61, 62, 63, 64].

2.2.1 Channel Correlation in Massive MIMO

The majority of the previously mentioned massive MIMO works have achieved the so-called channel hardening and favorable propagation properties by considering idealistic scenarios, in which the propagation paths were modeled to be uncorrelated, i.e., independent and identically distributed (i.i.d.) Rayleigh fading channels were considered. However, as many works have verified by field measurements [65, 66, 67], this scenario is hard to be obtained in real deployments. In practice, the scattering environment and insufficient inter-antenna space separation can make the channels to be strongly correlated, which has a non-negligible impact on the channel hardening and favorable propagation [68]. To be more precise, if the antenna elements are placed too close together, with high probability, the signals transmitted from adjacent antennas will propagate through similar paths. Consequently, two or more propagation channels will share common scatterers, and the spatial correlation will increase [69]. Furthermore, also due to the small antenna separation, the electromagnetic field generated by one antenna can change the current distribution of others, creating the effect known as mutual coupling. As stated in [70], the mutual coupling effect is unavoidable in co-located massive antenna arrays, and it is another factor that increases the channel correlation.

In conventional single-user MIMO systems, in order to achieve spatial multiplexing or diversity gains, the receiver needs to separate the signals transmitted by each BS antenna. Although, in highly correlated channels, this task can become quite complex, or even impractical. For this reason, for a long time, channel correlation has been considered as being detrimental to the performance of multi-antenna communication systems. However, when it comes to multi-user massive MIMO, the situation becomes more complex. The fact is that, depending on the scenario, correlation can be either beneficial or harmful to the system performance [65, 67, 68].

Figure 3: Massive MIMO system. Illustration of scenarios for achieving favorable propagation.

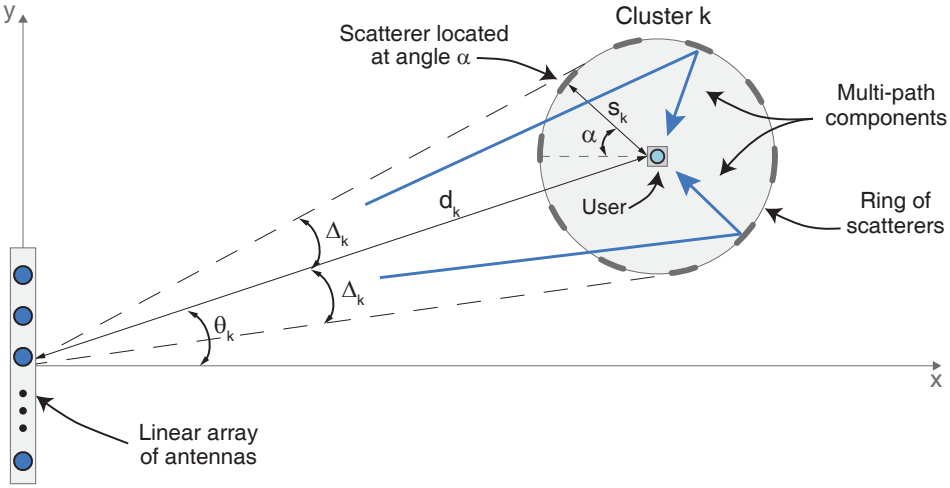


Source: Created by the author.

In order to understand the possible effects of correlation in massive MIMO, next, we discuss the statistical properties of the system channel responses.

First, let $\mathbf{h}_u \in \mathbb{C}^{M\times 1}$ be the correlated complex channel response between the BS and the user u. Under this definition, the transmit spatial channel covariance matrix of user u can be obtained as $\mathbf{R}_u = \mathbb{E}[\mathbf{h}_u \mathbf{h}_u^H] \in \mathbb{C}^{M\times M}$, which has rank denoted by r. In simple terms, $\mathbf{R}_u$ is a positive semi-definite Hermitian matrix that informs how the transmit antennas are correlated with each other. In uncorrelated scenarios, $\mathbf{R}_u$ is equal to the identity matrix $\mathbf{I}_M$, which is full-rank, i.e., $r = M$. However, in highly correlated environments, due to redundant propagation paths, the resulting covariance matrices can exhibit a low-rank behavior, in which only a small portion of eigenvalues is nonzero. This implies that the radiated energy will, most likely, propagate through a reduced number of directions, corresponding to the dominant nonzero eigenmodes. In [68], it has been demonstrated that rank-deficient covariance matrices can impact both channel hardening and favorable propagation properties. In particular, more antennas are necessary to reach (asymptotically) channel hardening in rank-deficient channels, making correlation detrimental for this property. On the other hand, depending on the spatial location of the users, favorable propagation can benefit or not from low-rank transmissions. For example, it can be shown that if users are receiving multi-path components from common clusters of scatterers and are located at similar azimuth angles, the probability of achieving mutual orthogonal channels is extremely low. In contrast, if users share different scattering clusters and are located at different angles, the chances of reaching the favorable propagation condition are high [67, 68]. Figure 3 illustrates both situations.

Figure 4: Illustration of the geometrical one-ring scattering model.



Source: Created by the author.

2.2.2 Spatial Correlation Model

In the last subsection, it became clear that correlation is an important parameter that must be taken into consideration while designing a multi-antenna wireless system. Therefore, in this dissertation, aiming to achieve accurate and realistic results, we consider scenarios with highly correlated channels. Specifically, for modeling the spatial correlation of propagation channels, we use the widely adopted one-ring scattering model [18, 71, 72]. Basically, it consists of a ray-tracing stochastic geometrical model, which assumes that clusters of scatterers are uniformly distributed on a ring that surrounds one or more users, as shown in Figure 4. This assumption is particularly valid for scenarios where the BS is elevated and unobstructed, such as in a communication tower, while users are located at ground level surrounded by many local scatterers like buildings, cars, trees, etc. One of the advantages of the one-ring model is that it makes it possible to determine inter-antenna correlation by exploring only physical properties of the system, such as antenna separation, array structure, angular spread, and azimuth angle. Besides, despite its simplicity, the model captures essential physical characteristics of the channels, which can be utilized for link-level simulations or for planing detailed field measurements [71]. In general, the following considerations are made in the one-ring model [73].

- Each real scatterer that lies in a certain angle α is represented by a corresponding virtual scatterer located at the same angle on the ring of scatterers.
- There is no line-of-sight (LOS) propagation.
- The virtual scatterers are assumed to be uniformly distributed in α.
- Only multi-path components that are reflected exactly once are considered.
- All waves impinging on the receive antennas are equal in power.

Under the above assumptions, and considering the existence of K spatial clusters, the (m,p) entry of the spatial channel covariance matrix for cluster k, denoted by $\mathbf{R}_k$, which represents the correlation between channels of antenna elements m and p, for $m,p = 1,\cdots,M$,

can be obtained by solving the following integral [71, 72]

$$[\mathbf{R}_k]_{m,p} = \frac{1}{2\Delta_k} \int_{-\Delta_k}^{\Delta_k} e^{j\mathbf{k}^T(\alpha+\theta_k)(\mathbf{u}_m-\mathbf{u}_p)} d\alpha, \tag{1}$$

where $\Delta_k \approx \tan^{-1}(\frac{s_k}{d_k})$ is the angular spread for the k-th cluster, which is located at a distance d_k and azimuth angle θ_k, surrounded by a ring of scatterers of radios s_k. $\mathbf{k}(\alpha) = -\frac{2\pi}{\lambda}[\cos(\alpha), \sin(\alpha)]^T$ is a vector that describes the incident planar waves arriving at the BS with angle of arrival α, and $\mathbf{u}_m, \mathbf{u}_p \in \mathbb{R}^{2\times 1}$ are coordinate vectors that specify the positions of transmit antennas m and p, respectively. For convenience, throughout this work, we suppose that $d_1 = d_2 = \cdots = d_K$ and $s_1 = s_2 = \cdots = s_K$.

Once the channel covariance matrix has been acquired, the correlated full channel matrix can be generated through the Karhunen-Loeve representation [72, 68]. The referred representation can be explained as follows. Let the eigenvalue decomposition of the channel covariance matrix be denoted by $\mathbf{R}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^H$, in which $\mathbf{\Lambda}_k$ is a $r^\star \times r^\star$ diagonal matrix containing the $r^\star$ nonzero eigenvalues of $\mathbf{R}_k$, and $\mathbf{U}_k \in \mathbb{C}^{M\times r^\star}$ are the corresponding eigenvectors. Then, by considering the existence of U users within the cluster k, the channel matrix for the u-th user can be decomposed as

$$\mathbf{H}_{k,u} = \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,u} \in \mathbb{C}^{M\times N}, \tag{2}$$

in which $\mathbf{G}_{k,u} \in \mathbb{C}^{r^\star \times N}$ represents a lower dimension fast-varying channel matrix.

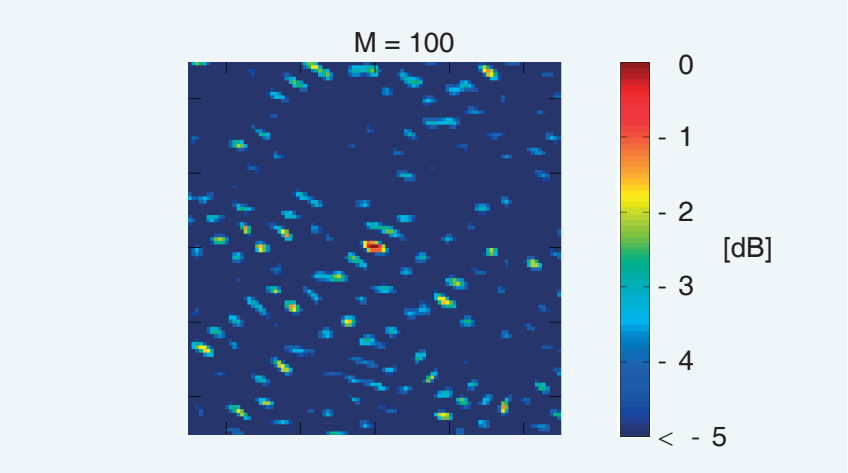
In this work, the above Karhunen-Loeve representation is employed for modeling the correlated wireless channels considered in our proposed systems.

2.2.3 Precoding for Massive MIMO

Precoding is a signal processing technique employed in MIMO systems to multiplex multiple users in space. It is of a common belief that precoding techniques (also known as beamforming) direct the signal propagation towards the receivers, forming the so-called beams. This is particularly true for LOS communication, where precoding enables the BS to concentrate the radiated energy into narrow beams towards the location of intended users. Although, in non-line-of-sight (NLOS) scenarios, it is not possible to steer the propagation through a direct path. Therefore, precoding in NLOS deployments, which is the case considered in this work, does not necessarily form beams, but rather adds up constructively the wavefronts in the desired point in space and destructively almost everywhere else [74]. This behavior can be visualized in Figure 5, where a massive MIMO system employing matched-filter precoding is transmitting to a user located in the center of the area displayed in the figure.

A vast number of precoding techniques are available in the literature, in which many different contexts and scenarios are contemplated. In particular, we draw our attention to a technique known as two-stage precoding [18, 7, 72], which is suitable for the scenarios considered

Figure 5: Energy propagated in area centered around the receiver. The figure shows the field strength generated with matched-filter precoding when $M = 100$.



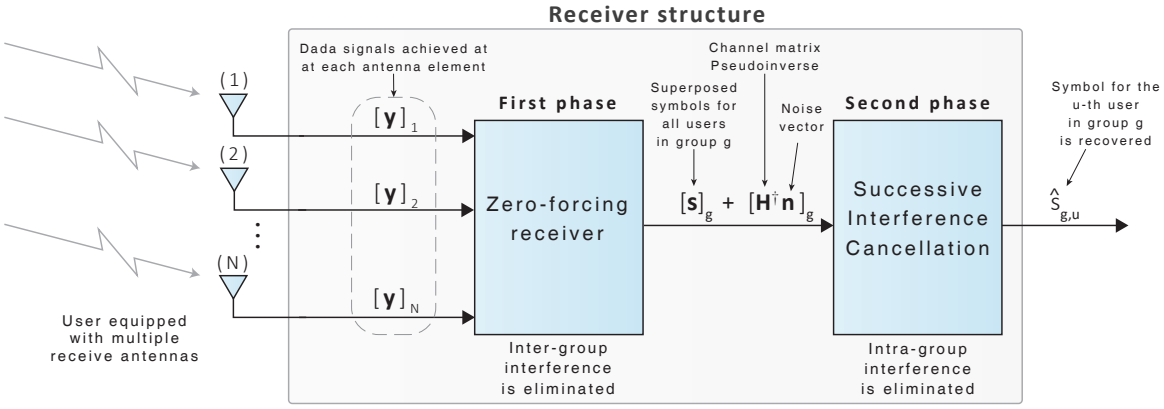
Source: Figure from [74].

in this dissertation. A two-stage precoder is composed of an outer preprocessing matrix $\mathbf{B}_k$ and an inner precoding vector $\mathbf{v}_{k,u}$. The matrix $\mathbf{B}_k$ is designed to focus the signal transmission to a desired spatial cluster k, such that everywhere else outside the area of interest the propagation is nulled out. To achieve this spatial directivity, the outer preprocessing matrix is constructed based on the null space spanned by the nonzero eigenmodes of the channel covariance matrices of interfering clusters, which can be modeled by the one-ring model presented in the last subsection. On its turn, the inner precoder vector, conventionally, is responsible for separating the users within the cluster, but it can be constructed in different ways, which can depend or not on the fast-varying channel matrices $\mathbf{G}_{k,u}$ in (2). The choice for the inner precoder will depend on the system objectives and the considered scenario. Specifically, in this dissertation, we project the precoders not to depend on the fast-varying channel matrices. This strategy provides some attractive advantages to massive MIMO setups, such as less processing overload and reduced feedback overhead. Detailed explanation for the construction of two-stage precoders is provided in Chapters 3 and 4.

2.2.4 NOMA in Massive MIMO

In Subsection 2.2.1, it was shown that if multiple users are located within the same cluster of scatterers, they cannot enjoy the favorable propagation property that can be achieved in massive MIMO. As a consequence, if the system is not designed with sophisticated precoding and detection strategies, these users can experience significant performance degradation. A classic solution for this issue can be obtained by employing the conventional TDMA technique to schedule the users within the cluster [6]. In this approach, the massive MIMO system serves only one user from each cluster at a time. The advantage of the strategy is that it enables each user to, separately, experience the beneficial favorable propagation phenomenon. However, as we have previously discussed, OMA techniques like TDMA are spectrally inefficient, reduce

Figure 6: Simplified diagram for a multi-antenna MIMO-NOMA receiver.



Source: Created by the author.

the connectivity capacity, and provide increased latency. Fortunately, NOMA can be efficiently combined with massive MIMO to tackle the difficulties faced in scenarios with highly correlated channels. It has been demonstrated that NOMA can be very beneficial to MIMO systems in environments where various users share a common cluster of scatterers [7, 44, 75]. The integration of NOMA and massive MIMO enables the BS to employ a single precoder to serve all users within a given spatial cluster. In particular, by employing two-stage precoders at the BS, the inter-cluster interferences can be effectively eliminated. Then, NOMA can be employed by the users within each cluster to cancel the remaining intra-cluster interference.

Theoretically, there is no limit to the number of users that can be served with NOMA. However, in reality, when the number of users within the cluster rises, the SIC complexity increases, and the reception performance is impacted. To overcome this impairment, some works have come up with the idea of sub-dividing the spatial clusters into multiple groups with fewer users [7, 6, 43]. In this approach, instead of canceling the interference of everyone in each user within the cluster, the SIC process is distributed among the various smaller groups. To be more specific, each user will perform SIC by considering only the interfering messages of users within its own group. As a result, the system complexity will be significantly reduced. This strategy can be accomplished by dividing the intra-cluster interference cancellation into two phases at the users' side. In the first phase, in order to mitigate the inter-group interference, the users employ some detection technique, such as maximum likelihood (ML), minimum mean squared error (MMSE) or zero-forcing. As output of this first phase, each user will obtain a superposed symbol containing the messages of all other users within the group. Then, in the second phase, each user performs SIC to cancel intra-group interference and finally recover the desired message. This process can be visualized in Figure 6, in which the structure of a multi-antenna receiver is illustrated. Due to the inherent low-complexity of zero-forcing receivers, they will be adopted as standard in the designs proposed in this dissertation.

An important and worth mentioning peculiarity of massive MIMO-NOMA systems is that SC and SIC is not carried out in the same way as in single-antenna systems. As ex-

emplified in Subsection 2.1.3, for single-antenna systems, the power allocation at the BS and the SIC decoding order at the receivers are conducted based on the actual scalar channel gains. However, the users in massive MIMO-NOMA systems do not have a unique scalar channel. Instead, they have a matrix of channel coefficients, which makes the application of NOMA a bit more complicated. To cope with such challenging characteristic, the users are instead ordered based on what is called effective channel gains, which are the result of the multiplication of precoding, channel, and detection matrices. For example, assuming that a two-stage precoder is employed in a multi-cluster multi-group massive MIMO-NOMA scenario, where the channel matrix for u-th user in the g-th group in the k-th cluster is denoted by $\mathbf{H}_{k,g,u}$, and assuming that the users employs for reception a generic detection matrix denoted as $\mathbf{D}_{k,g,u}$, then, the effective channel gains can be defined[1] by $\gamma_{k,g,u} = |\mathbf{D}_{k,g,u}\mathbf{H}_{k,g,u}^H\mathbf{B}_k\mathbf{p}_{k,g}|^2$, so the users of this system would be ordered based on the values of $\gamma_{k,g,u}$.

2.2.5 Zero-Forcing Receiver

Zero-forcing receivers in massive MIMO systems follow the same principles of conventional zero-forcing equalizers employed in frequency-selective channels [23, 73]. The great attractiveness of these receivers is that by applying only a simple linear transformation, the effects of the fading channels are perfectly reversed. Therefore, the strategy described in last subsection of employing a zero-forcing receiver in MIMO-NOMA users, in addition to the low-computational complexity advantage, it completely eliminates the inter-group interference. To better illustrate how the reception process is performed, consider the following example.

Suppose that a vector of superposed symbols, denoted by $\mathbf{s} = [s_1, s_2, \cdots, s_G]^T$, has been transmitted through fast-fading channels to G NOMA groups, in which the g-th element of $\mathbf{s}$ is intended to the g-th group. In addition, let $\bar{\mathbf{H}} \in \mathbb{C}^{N\times G}$ be the channel matrix after being multiplied by the precoding matrices (known as virtual channel), where N represents the number of receive antennas for each user. Then, the output of the zero-forcing receiver can be written as

$$\hat{\mathbf{s}} = [s_1, s_2, \cdots, s_G]^T + \mathbf{H}^{\dagger}\mathbf{n}, \tag{3}$$

where $\mathbf{n}$ is a noise vector, and $\mathbf{H}^{\dagger}$ is the zero-forcing linear transformation matrix responsible for reversing the channel effects. If the channel matrix $\bar{\mathbf{H}} \in \mathbb{C}^{N\times G}$ satisfies $G \leq N$, then the transformation matrix can be given by

$$\mathbf{H}^{\dagger} = (\bar{\mathbf{H}}^H\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}^H, \tag{4}$$

which is known as the left Moore-Penrose pseudo-inverse of $\bar{\mathbf{H}}$.

[1] Depending on the detection and precoding techniques, different effective gain expressions can be achieved, i.e., different designs will result in different effective channel gains.

As can be observed in (3), the zero-forcing receiver has decoupled the received signal into G parallel symbols, in which the inter-group interference has been completely eliminated. This interesting feature enables the users of each group to apply SIC to their corresponding superposed symbol in $\hat{\mathbf{s}}$ and recover the desired data without any disruption from interfering groups. As a consequence, the computational complexity imposed by SIC can be efficiently alleviated. Nevertheless, even though the superposed symbols for different groups have been effectively separated, the noise component is being amplified by the channel pseudo-inverse. This characteristic can degrade the system performance in deployments with high noise levels. Besides, the inversion of bad-conditioned channel matrices can also contribute to noise enhancement, making strongly correlated scenarios detrimental do the performance of zero-forcing receivers [76, 77].

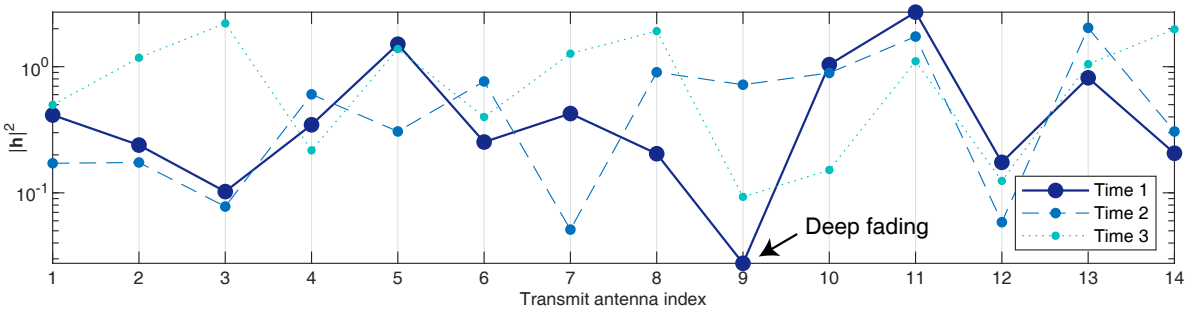
2.3 Concepts of Multi-Polarization

In this section, we introduce the basic concepts of multi-polarization necessary for the complete understanding of Chapter 3. In simple words, polarization is defined as the dominant direction of the radiated electric field vector with respect to the horizontal plane, i.e., with respect to the earth's surface. The most common polarization choices employed in commercial cellular systems are 0° (horizontal polarization), 90° (vertical polarization), and $\pm 45^\circ$ (slant polarization) [68, 73]. Ideally, the electromagnetic waves transmitted by orthogonally polarized antennas[2], e.g. 0° and 90°, or $+45^\circ$ and -45°, experience uncorrelated fading channels, independently of the inter-antenna space separation. Therefore, it is possible to install two antenna elements with orthogonal polarizations in the same physical location at both transmit and receive sides without generating any cross-polar interference. As a result, the employment of dual-polarized antenna arrays can double the degrees of freedom, i.e., the number of antennas, of MIMO with the advantage of occupying the same physical space of conventional single-polarized arrays. In addition to the space efficiency, it has been shown that massive MIMO systems employing dual-polarization can outperform the single-polarized counterpart with the same number of antenna elements [18, 21, 22]. With these benefits, multi-polarization in massive MIMO has become the standard choice for LTE-A deployments, and it is expected to be part of 5G and beyond [29, 78, 79].

The assumption of zero cross-polar interference is valid for idealistic scenarios where receive antennas with a certain polarization are capable of detecting signals only from transmit antennas with their identical corresponding polarization. Although, in real-world deployments, due to two depolarization phenomena, the dual-polarized receivers will suffer from non-negligible cross-polar interference. The first cause of depolarization is known as finite (or imperfect) antenna cross-polar isolation (XPI), and the second is named as cross-polar ratio (XPR). The XPI property arises from the impossibility of building perfectly polarized antennas,

[2] In this work, the dual-polarized massive MIMO-NOMA systems are designed with co-located horizontally and vertically polarized antennas.

Figure 7: Channel gains versus transmit antennas in different instants of time.



Source: Created by the author.

meaning that, the electric field generated by practical antennas exhibits a non-zero cross-polar radiation pattern. On the other hand, the XPR is a scatterer-based depolarization phenomenon, in which the effects of reflection, diffraction, and diffuse scattering introduce power imbalance between the orthogonal polarizations [68, 73]. These two phenomena can impact the system performance. Therefore, it is important to consider the depolarization effects when modeling polarization in MIMO channels.

The mentioned depolarization phenomena can be modeled through the so-called cross-polar discrimination (XPD), which accounts for the combined impact of both XPI and XPR [73]. By considering that the antennas have very low cross-polar radiation patterns, the XPI factor can be neglected[3] [18, 68], that is, it can be assumed that XPD = XPR. Under this assumption, only the power imbalance caused by the scattering medium needs to be characterized. This task can be accomplished through the following matrix [18, 19, 68, 73]

$$\boldsymbol{\chi} = \begin{bmatrix} 1 & \sqrt{\chi} \\ \sqrt{\chi} & 1 \end{bmatrix}, \tag{5}$$

where χ stands for the inverse of the XPD, in which $0 \leq \chi \leq 1$.

The parameter χ models the energy leakage between orthogonal polarizations, in which $\chi = 0$ represents the ideal case of zero cross-polar interference, and $\chi = 1$ corresponds to the extreme scenario with the highest possible cross-polar interference. The matrix $\boldsymbol{\chi}$ is used for modeling the dual-polarized channels in Chapter 3, where further details are provided.

2.4 Concepts of Diversity

In this subsection, we introduce some basic concepts of diversity, which is important for the contributions developed in both chapters 3 and 4. As mentioned in the Introduction Section, even with all the advantages that massive MIMO-NOMA can offer, in many situations,

[3]Throughout Chapter 3, it is considered that the depolarization effects are caused exclusively by the scattering environment, i.e., the XPI phenomenon is neglected.

the system performance can still be degraded. Diversity techniques[4] are one of the most effective low-complexity solutions for overcoming the impairments of fading channels, and also for mitigating other possible degradation causes [23, 73]. The basic idea behind any diversity technique is to send multiple replicas of the same information over different propagation paths. These multiple signal copies, known as diversity branches, must arrive at the receivers through uncorrelated fading channels, so the chances of receiving the desired information with high reliability are maximized. Once the multiple replicas have been received, by applying some diversity combining technique, such as selection combining, maximal-ratio combining, and others, the system performance can be significantly improved.

Diversity can be obtained from a variety of dimensions, such as polarization, space, and time. For instance, systems equipped with dual-polarized arrays exhibit a low correlation between polarizations. Therefore, each polarization can be treated as a diversity branch, which can be combined to improve communication performance. The main limitation of polarization diversity in dual-polarized systems is that it can only offer two diversity branches. Spatial diversity can be realized in point-to-point MIMO systems when the antennas are placed far enough apart, i.e., separated by more than half of the wavelength, so spatial correlation and mutual coupling effects can be avoided. In these multi-antenna systems, more than two diversity branches can be obtained, but physical space constraints can limit their application [80]. While time diversity techniques explore the temporal fluctuation of the channel to exploit diversity, in which redundant data symbols are distributed over different slots of time. However, in order to time diversity be effective, the separation between time-slots must be greater than the coherence time of the channel [23], which makes the number of branches be limited by the maximum delay tolerable in the system. Figure 7 illustrates the channel fluctuations of a multi-antenna system in different coherent times. One can observe that the multiple diversity branches help to overcome the deep fading faced by some individual channels.

In Chapter 3, we make use of the polarization domain to enhance the performance of our multi-polarized design, and, in Chapter 4, the space and time dimensions are exploited to propose a new diversity scheme.

[4] Diversity can be combined with sophisticated coding techniques to efficiently trade-off between multiplexing and diversity gains. However, coding goes beyond the scope of this work.

3 MASSIVE MIMO-NOMA NETWORKS WITH MULTI-POLARIZED ANTENNAS

In this chapter, the benefits of multi-polarization in multi-cluster multi-user massive MIMO-NOMA networks are investigated. Firstly, motivations and main contributions are presented in Section 3.1. Then, in Section 3.2, the system model and important considerations for our analysis are introduced. In Section 3.3, the precoder designs and the respective reception strategies are proposed. Following, in Section 3.4, an in-depth analytical analysis is carried out. In Section 3.5, numerical and simulation results are presented, while Section 3.6 brings the summary of the chapter.

3.1 Motivations and Contributions

It has been shown that all the advantages achieved in massive MIMO are amplified when the number of transmit antenna grows. However, if the antenna elements are not separated by at least half of the wavelength, the system performance is severely degraded. Due to this impairing characteristic, in practical massive MIMO implementations, the maximum number of antennas is limited. This issue is even more critical at the users' side since, usually, the users' devices are very compact. Fortunately, as mentioned in Section 1.1, the use of dual-polarized antennas can effectively alleviate the space constraints at both BS and users' terminals. Although, to the best of the author's knowledge, a full and in-depth understanding of the combination of multi-polarization and massive MIMO-NOMA systems has not been examined in the literature so far. Therefore, motivated by this lacuna, this chapter aims to design and analyze the performance of multi-cluster multi-user dual-polarized massive MIMO-NOMA systems. In particular, the downlink mode is assumed, in which a single base station communicates with multiple users, with all terminals being equipped with multiple co-located dual-polarized antennas. Specific details and main contributions of this chapter can be summarized as follows:

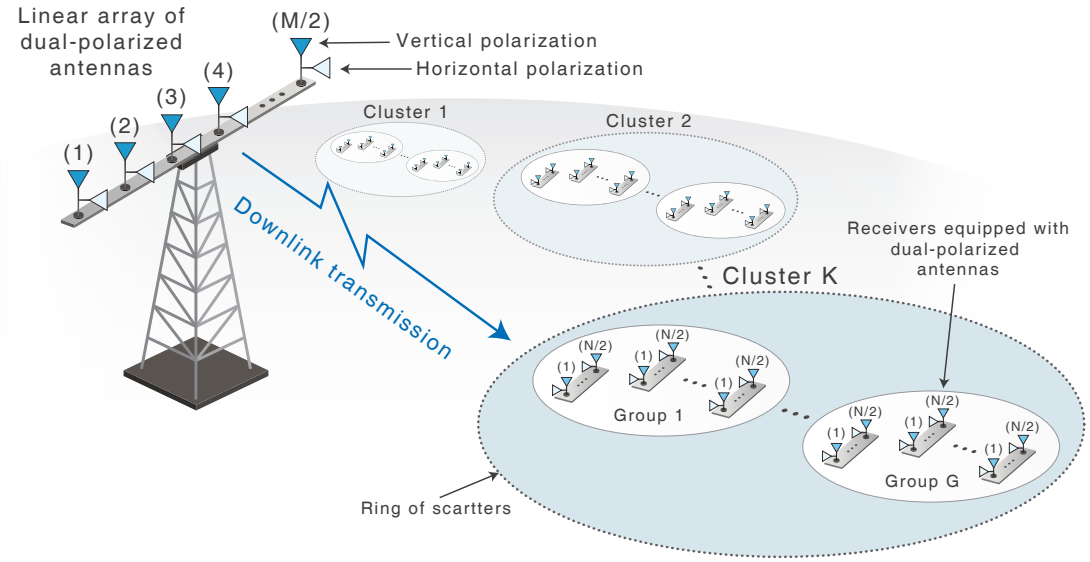
- We adopt a dual-stage precoder that is designed based only on the slowly varying spatial correlation information of the channel. The inner precoder is built to provide either multiplexing or diversity gains through the polarization domain.
- Based on the inner precoder choice, two approaches with two different reception strategies are proposed for the considered dual-polarized massive MIMO-NOMA system. Specifically, assuming a moderate to high cross-polar transmission interference, the first approach maximizes the number of user groups that are simultaneously served. On the other hand, considering a low cross-polar interference scenario, the second approach is proposed, and it improves the system performance through polarization diversity, at the expense of reducing the maximum number of supported groups.
- An in-depth analysis is performed, in which we obtain a novel analytical framework that provides a simpler and practical alternative for evaluation and design of efficient dual-polarized massive MIMO-NOMA systems. In particular, closed-form expressions for the outage probability are derived for both approaches, based on which the respective

asymptotic studies are carried out and the diversity gains are determined. The ergodic sum-rates are also derived.

- Representative numerical examples are presented along with insightful discussions. For instance, our results show that the proposed dual-polarized MIMO-NOMA designs outperform conventional single-polarized systems, even for high cross-polar interference.
- Simulation results are plotted to corroborate the analytical framework and analysis.

3.2 System Model

Figure 8: System model. Both BS and users are equipped with multiple co-located dual-polarized antennas.



Source: Created by the author.

Let a downlink transmit scenario in which a single BS communicates with multiple users, as illustrated in Figure 8. The BS and the users are equipped, respectively, with $M/2$ and $N/2$ pairs of co-located dual-polarized antennas, with $M \gg N$. Due to the dual-polarized structure, the total number of transmit and receive antennas are considered to be even. Furthermore, the users are assumed to be surrounded by local scatterers, forming K spatial clusters that follow the geometrical one-ring scattering model of Subsection 2.2.2. Within each cluster k, there are L users that share the same spatial covariance matrix $\mathbf{R}_k \in \mathbb{C}^{M \times M}$. Besides, since each pair of polarized antennas are co-located, we assume equal covariance matrices for both polarizations.

With the given design and considerations, the correlated dual-polarized full channel matrix for the l-th user at the k-th cluster can be decomposed and expressed in the following

block structure [18, 19]

$$\begin{aligned}\mathbf{H}_{k,l} &= \left(\mathbf{I}_2 \otimes \left(\mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}}\right)\right)\left(\begin{bmatrix}\mathbf{G}_{k,l}^{v\to v} & \mathbf{G}_{k,l}^{h\to v}\\ \mathbf{G}_{k,l}^{v\to h} & \mathbf{G}_{k,l}^{h\to h}\end{bmatrix} \odot \left(\boldsymbol{\chi} \otimes \mathbf{1}_{r_k^\star \times \frac{N}{2}}\right)\right)\\ &= \begin{bmatrix}\mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{v\to v} & \sqrt{\chi}\mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{h\to v}\\ \sqrt{\chi}\mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{v\to h} & \mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{h\to h}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{H}_{k,l}^{v\to v} & \mathbf{H}_{k,l}^{h\to v}\\ \mathbf{H}_{k,l}^{v\to h} & \mathbf{H}_{k,l}^{h\to h}\end{bmatrix} \in \mathbb{C}^{M\times N},\end{aligned} \tag{6}$$

with covariance matrix given by

$$\mathbf{R}_k = (\chi + 1)\begin{bmatrix}\tilde{\mathbf{R}}_k & \mathbf{0}\\ \mathbf{0} & \tilde{\mathbf{R}}_k\end{bmatrix}, \tag{7}$$

where $\chi \in [0,1]$ stands for the inverse of the antennas' cross-polar discrimination parameter (XPD), explained in Section 2.3, $\tilde{\mathbf{R}}_k \in \mathbb{C}^{\frac{M}{2}\times\frac{M}{2}}$ represents the covariance matrix of rank r_k, corresponding to one polarization, $\boldsymbol{\Lambda}_k$ is the $r_k^\star \times r_k^\star$ diagonal matrix containing $r_k^\star$ nonzero eigenvalues of $\tilde{\mathbf{R}}_k$, with elements ordered in decreasing order, $\mathbf{U}_k \in \mathbb{C}^{\frac{M}{2}\times r_k^\star}$ denotes the matrix composed of eigenvectors of $\tilde{\mathbf{R}}_k$ corresponding to the nonzero eigenvalues, and $\mathbf{G}_{k,l}^{i\to j} \in \mathbb{C}^{r_k^\star \times \frac{N}{2}}$ is the fast-varying channel matrix of the link between the i-th polarized transmit antenna and the j-th polarized receive antenna, $i,j \in \{v,h\}$, whose elements follow the complex Gaussian distribution with zero mean and unit variance. The letters h and v are used to indicate horizontal and vertical polarizations, respectively.

In order to enable the implementation of NOMA, following the strategy discussed in Subsection 2.2.4, the L users within a cluster are subdivided into G smaller groups of U users each, such that $L = UG$. Under this assumption, the superposition coding technique is applied to the data intended for users within a group. Thus, the transmitted data signal can be written as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{B}_k \sum_{g=1}^{G} \mathbf{p}_{k,g} \sum_{u=1}^{U} \alpha_{k.g,u} s_{k,g,u}, \tag{8}$$

where $s_{k,g,u}$ and $\alpha_{k,g,u}$ are, respectively, the data symbol and the power allocation coefficient for the u-th user in the g-th group of the k-th cluster. Note that $\sum_{u=1}^{U} \alpha_{k.g,u}^2 = 1$ so that the total transmission power allocated to each NOMA group is normalized to 1. $\mathbf{B}_k \in \mathbb{C}^{M\times\bar{M}}$ is the preprocessing matrix which is designed based on the long-term channel covariance matrix $\mathbf{R}_k$, with $\bar{M}$ being a design parameter which defines the effective number of data streams that arrives at the receiver, and $\mathbf{p}_{k,g} \in \mathbb{C}^{\bar{M}\times 1}$ is the precoding vector for the g-th group.

3.3 Precoder Designs and Signal Reception

As introduced in Subsection 2.2.3, the outer preprocessing matrix $\mathbf{B}_k$ has the role of eliminating the interferences from other clusters, while the inner precoding vector $\mathbf{p}_{k,g}$ can be construct in different ways according to the objectives of the system. In this dissertation, we choose to design $\mathbf{p}_{k,g}$ only to assign the superimposed data symbols to each group within a cluster. More details about $\mathbf{p}_{k,g}$ will be provided later.

Firstly, let us focus on the design of the preprocessing matrix $\mathbf{B}_k$. Our purpose is to extinguish the interfering signals coming from other clusters. Mathematically speaking, we must have

$$\mathbf{H}_{k,l}^H \mathbf{B}_{k'} \approx \mathbf{0}, \forall k \neq k'. \tag{9}$$

To accomplish this goal, the subspace spanned by the preprocessing matrix must be near-orthogonal (perfect orthogonality is achieved when $r_k^\star = r_k$) to the subspace spanned by the dominant eigenvectors of other clusters [72], which means that $\mathbf{B}_k$ must be built based on the null space of the dominant eigenmodes of interfering clusters. Then, we assume that $r_1^\star = r_2^\star = \ldots = r_K^\star = r^\star$, $r_1 = r_2 = \cdots = r_K = r$ and we define the following matrix

$$\mathbf{U}_k^- = [\mathbf{U}_1, \cdots, \mathbf{U}_{k-1}, \mathbf{U}_{k+1}, \cdots, \mathbf{U}_K] \in \mathbb{C}^{\frac{M}{2} \times (K-1)r^\star}. \tag{10}$$

Next, we apply the singular value decomposition (SVD) in $\mathbf{U}_k^-$ and express its left eigenvectors as $\mathbf{E}_k = [\mathbf{E}_k^{(1)}, \mathbf{E}_k^{(0)}]$, where $\mathbf{E}_k^{(0)} \in \mathbb{C}^{\frac{M}{2} \times \frac{M}{2} - (K-1)r^\star}$ represents the eigenvectors that correspond to the last $\frac{M}{2} - (K-1)r^\star$ vanishing eigenvalues of $\mathbf{U}_k^-$. Since $(\mathbf{E}_k^{(0)})^H \mathbf{U}_k^- = \mathbf{0}$, the projected channel that is orthogonal to the dominant $r^\star$ eigenmodes of the groups $k' \neq k$ can be defined as $\hat{\mathbf{H}}_{k,l} = (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)})^H \mathbf{H}_{k,l}$. Consequently, the covariance matrix of $\hat{\mathbf{H}}_{k,l}$ is given by

$$\hat{\mathbf{R}}_k = (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)})^H \mathbf{R}_k (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)}), \tag{11}$$

and the covariance matrix considering only one of the polarizations can be written as

$$\underline{\hat{\mathbf{R}}}_k = (\mathbf{E}_k^{(0)})^H \tilde{\mathbf{R}}_k \mathbf{E}_k^{(0)}. \tag{12}$$

By defining the left eigenvectors of $\underline{\hat{\mathbf{R}}}_k$ as $\mathbf{F}_k = [\mathbf{F}_k^{(1)}, \mathbf{F}_k^{(0)}]$, in which $\mathbf{F}_k^{(1)} \in \mathbb{C}^{\frac{M}{2} - (K-1)r^\star \times \frac{\bar{M}}{2}}$ consists of the first $\frac{\bar{M}}{2}$ columns of $\mathbf{F}_k$, the precoder design can be obtained as follows

$$\tilde{\mathbf{B}}_k = \mathbf{E}_k^{(0)} \mathbf{F}_k^{(1)} \in \mathbb{C}^{\frac{M}{2} \times \frac{\bar{M}}{2}}, \tag{13}$$

$$\mathbf{B}_k = \begin{bmatrix} \tilde{\mathbf{B}}_k & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{B}}_k \end{bmatrix} \in \mathbb{C}^{M \times \bar{M}}, \tag{14}$$

in which the following constraints must hold

$$K \leq \bar{M} \leq 2\left(\frac{M}{2} - (K-1)r^{\star}\right), \tag{15}$$

and

$$\bar{M} \leq 2r^{\star} \leq 2r. \tag{16}$$

It is important to highlight that $r^{\star}$ is a design parameter that should be adjusted based on how much inter-cluster interference is tolerable in the system. The closer the value of $r^{\star}$ gets to r, less inter-cluster interference the system will experience. However, as stated in [72], choosing $r^{\star}$ too large, e.g., $r^{\star} = r$, does not bring significant improvements. Instead, it will only increase the factor $(K-1)r^{\star}$ and, since the constraint $(K-1)r^{\star} < \frac{M}{2}$ must be obeyed, this will eventually force a reduction on the number of clusters that are simultaneously served. Therefore, the parameter $r^{\star}$ should be carefully chosen in order to attend the system requirements. In our implementation, given the constraints in (15) and (16), we choose to configure the dominant eigenvalues as $r^{\star} = \min\left\{r, \left\lfloor (\frac{M}{2} - \frac{\bar{M}}{2})\frac{1}{K-1} \right\rfloor\right\}$.

Since it is not an easy task to acquire the fast-fading matrices at the transmitter and in order to reduce the feedback overhead, it is assumed that the BS has access only to the slow-fading covariance matrix $\tilde{\mathbf{R}}_k$. Thus, the precoding vector $\mathbf{p}_{k,g}$ is designed in a way that it will not depend on the short-term channel state information. As aforementioned, the only role of $\mathbf{p}_{k,g}$ is to assign the superimposed data symbols to its respective group.

3.3.1 Precoder for Polarization Multiplexing - Approach I

In this first precoder design choice, we aim to maximize the number of groups that are simultaneously served within a cluster. To this end, we choose to assign to each group a different data stream so that a maximum of $\bar{M}$ parallel transmissions can be performed. Thus, supposing that $G \leq \bar{M}$, the precoding vector can be defined as

$$\mathbf{p}_{k,g} = [\ \mathbf{0}_{1\times(g-1)}\ ,\ 1\ ,\ \mathbf{0}_{1\times(\bar{M}-g)}\ ]^{T}. \tag{17}$$

Note that, with the above precoder[5], the g-th effective data stream is assigned to the g-th group.

[5] In order to provide insightful performance comparisons, similarly as in the dual-polarized system, the single-polarized implementations also employ two-stage precoders. The same inner precoder strategy of approach I is adopted, and the outer precoding matrix is constructed identically as in Section 3.3.

3.3.2 Precoder for Polarization Diversity - Approach II

In order to provide diversity gain through the polarization domain, the precoder is now designed to transmit the same data symbol, intended to the g-th group, in both polarizations. This approach limits to $\frac{\bar{M}}{2}$ the total number of groups that are simultaneously served in a cluster. Then, assuming that $G \leq \frac{\bar{M}}{2}$, the precoding vector is chosen as

$$\mathbf{p}_{k,g} = \begin{bmatrix} \mathbf{p}^{v}_{k,g} \\ \mathbf{p}^{h}_{k,g} \end{bmatrix} = \begin{bmatrix} [\, \mathbf{0}_{1\times(g-1)} \;,\; 1 \;,\; \mathbf{0}_{1\times(\frac{\bar{M}}{2}-g)} \,]^{T} \\ [\, \mathbf{0}_{1\times(g-1)} \;,\; 1 \;,\; \mathbf{0}_{1\times(\frac{\bar{M}}{2}-g)} \,]^{T} \end{bmatrix}, \tag{18}$$

where $\mathbf{p}^{p}_{k,g} \in \mathbb{C}^{\frac{\bar{M}}{2}\times 1}$ denotes the sub-precoding vector corresponding to the polarization $p \in \{v,h\}$. With this design, the g-th pair of effective data streams, one stream from each polarization, is selected to transmit to the g-th group.

It is noteworthy that the two above approaches do not introduce any additional changes in the preprocessed signal. Therefore, the task of eliminating inter-group interference is left entirely to the user's terminal.

3.3.3 Signal Reception

After the superimposed data signal is transmitted by the BS through the dual-polarized fast fading channels, the u-th user at the g-th group of the k-th cluster observes the following

$$\mathbf{y}_{k,g,u} = \begin{bmatrix} \mathbf{y}^{v}_{k,g,u} \\ \mathbf{y}^{h}_{k,g,u} \end{bmatrix} = \mathbf{H}^{H}_{k,g,u} \sum_{l=1}^{K} \mathbf{B}_{l} \sum_{m=1}^{G} \mathbf{p}_{l,m} \sum_{n=1}^{U} \alpha_{l,m,n} s_{l,m,n} + \mathbf{n}_{k,g,u}, \tag{19}$$

where $\mathbf{y}^{v}_{k,g,u} \in \mathbb{C}^{\frac{N}{2}\times 1}$ and $\mathbf{y}^{h}_{k,g,u} \in \mathbb{C}^{\frac{N}{2}\times 1}$ are, respectively, the information vectors received at the vertically and horizontally polarized antennas, and $\mathbf{n}_{k,g,u} \in \mathbb{C}^{N\times 1}$ is the noise vector with entries obeying a complex Gaussian distribution with zero-mean and variance σ_n^2.

Since we consider the system with a massive number of transmit antennas, in which $M \gg N$, and assume that the clusters have non-overlapping azimuth angles, for a reasonable value of $r^{\star}$, the near-orthogonality condition of (9) can be satisfied [18, 72]. Under this consideration, the residual interference that arrives from other clusters will be extremely small, i.e. it can be neglected. Thus, by making the assumption that $\mathbf{B}_k$ nulls out the inter-cluster interferences, we simplify the signal in (19) as

$$\mathbf{y}_{k,g,u} = \mathbf{H}^{H}_{k,g,u} \mathbf{B}_{k} \sum_{m=1}^{G} \mathbf{p}_{k,m} \sum_{n=1}^{U} \alpha_{k,m,n} s_{k,m,n} + \mathbf{n}_{k,g,u}. \tag{20}$$

Now, in order to eliminate the inter-group interference and recover the desired signal, the zero-forcing equalizer of Subsection 2.2.5 is adopted at the users side. Supposing that $N \geq \bar{M}$, the

Moore-Penrose pseudo-inverse of the virtual channel $\bar{\mathbf{H}}_{k,g,u} = \mathbf{H}_{k,g,u}^{H}\mathbf{B}_k \in \mathbb{C}^{N\times\bar{M}}$ can be defined as $\bar{\mathbf{H}}_{k,g,u}^{\dagger} = (\bar{\mathbf{H}}_{k,g,u}^{H}\bar{\mathbf{H}}_{k,g,u})^{-1}\bar{\mathbf{H}}_{k,g,u}^{H} \in \mathbb{C}^{\bar{M}\times N}$. After multiplying (20) by $\bar{\mathbf{H}}_{k,g,u}^{\dagger}$, the interference is removed and the u-th user acquires a vector formed by the noisy version of the transmitted superimposed symbols, i.e.,

$$\hat{\mathbf{s}}_{k,g,u} = \sum_{m=1}^{G} \mathbf{p}_{k,m} \sum_{n=1}^{U} \alpha_{k,m,n} s_{k,m,n} + \bar{\mathbf{H}}_{k,g,u}^{\dagger}\mathbf{n}_{k,g,u}. \tag{21}$$

The users retrieve their messages from the recovered superimposed data that was assigned to the group that they belong to. However, the messages recovery depend on the type of precoder that is employed at the BS, i.e., the precoding vector can be formulated either as (17) or as (18), and this is further detailed next.

Approach I: It is considered that the system is operating on multiplexing mode, that is, the precoder in (17) is employed at the BS, and, for convenience, we assume that the number of groups is $G = \bar{M}$. Under these assumptions, the detected vector in (21) can be rewritten as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \sum_{n=1}^{U} \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^{U} \alpha_{k,G,n} s_{k,G,n} \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^{\dagger}\mathbf{n}_{k,g,u}. \tag{22}$$

One can see that each element of (22) belongs to a specific group. Therefore, the u-th user at the g-th group recovers its message from the g-th element of the received vector $\hat{\mathbf{s}}_{k,g,u}$, which yields

$$\hat{s}_{k,g,u} = [\hat{\mathbf{s}}_{k,g,u}]_g = \sum_{n=1}^{U} \alpha_{k,g,n} s_{k,g,n} + [\bar{\mathbf{H}}_{k,g,u}^{\dagger}\mathbf{n}_{k,g,u}]_g. \tag{23}$$

It is worth mentioning that the main advantage of this strategy lies mainly in the high number of groups that can be simultaneously served, besides its simplicity. However, as it can be realized, each user recovers its desired information by exploring only one of the polarizations, either vertical or horizontal, what may impact the system performance.

Approach II: For this approach, the BS employs the precoding vector as in (18), which makes the system to operate in diversity mode. For simplicity, it is assumed that $G = \frac{\bar{M}}{2}$ groups are simultaneously served. Thus, the vector in (21) can be expressed as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \begin{bmatrix} \sum_{n=1}^{U} \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^{U} \alpha_{k,G,n} s_{k,G,n} \end{bmatrix} \\ \begin{bmatrix} \sum_{n=1}^{U} \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^{U} \alpha_{k,G,n} s_{k,G,n} \end{bmatrix} \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^{\dagger}\mathbf{n}_{k,g,u} = \begin{bmatrix} \mathbf{s}_k^{v} \\ \mathbf{s}_k^{h} \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^{\dagger}\mathbf{n}_{k,g,u}, \tag{24}$$

where $\mathbf{s}_k^v \in \mathbb{C}^{G\times 1}$ and $\mathbf{s}_k^h \in \mathbb{C}^{G\times 1}$ are, respectively, the data vectors transmitted by the vertically and horizontally polarized antennas intended to the k-th cluster. The g-th element of both $\mathbf{s}_k^v$ and $\mathbf{s}_k^h$ consists of superimposed data symbols that are intended to the g-th group. Differently from the first approach, now we take advantage of the polarization diversity. It is assumed that the inverse of the XPD is small, e.g, $\chi \ll 1$. With this assumption, the off-diagonal blocks of the channel matrix in (6) approach to zero, i.e., the cross-polar interference becomes negligible. As a result, the virtual channel matrix can be simplified in the following block diagonal structure

$$\bar{\mathbf{H}}_{k,g,u} = \begin{bmatrix} (\mathbf{H}_{k,g,u}^{v\to v})^H \tilde{\mathbf{B}}_k & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}_{k,g,u}^{h\to h})^H \tilde{\mathbf{B}}_k \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{v} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^{h} \end{bmatrix}. \tag{25}$$

Consequently, its corresponding zero-forcing detection matrix becomes

$$\begin{aligned} \bar{\mathbf{H}}_{k,g,u}^{\dagger} &= \begin{bmatrix} ((\bar{\mathbf{H}}_{k,g,u}^{v})^H \bar{\mathbf{H}}_{k,g,u}^{v})^{-1} (\bar{\mathbf{H}}_{k,g,u}^{v})^H & \mathbf{0} \\ \mathbf{0} & ((\bar{\mathbf{H}}_{k,g,u}^{h})^H \bar{\mathbf{H}}_{k,g,u}^{h})^{-1} (\bar{\mathbf{H}}_{k,g,u}^{h})^H \end{bmatrix} \\ &= \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \end{bmatrix}. \end{aligned} \tag{26}$$

From above, the signal in (24) can be rewritten as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{k,g,u}^{v} \\ \mathbf{y}_{k,g,u}^{h} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} \mathbf{y}_{k,g,u}^{v} \\ \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \mathbf{y}_{k,g,u}^{h} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{s}}_{k,g,u}^{v} \\ \hat{\mathbf{s}}_{k,g,u}^{h} \end{bmatrix}, \tag{27}$$

where $\bar{\mathbf{H}}_{k,g,u}^{\dagger v}$ and $\bar{\mathbf{H}}_{k,g,u}^{\dagger h}$ are the zero-forcing matrices corresponding to vertical and horizontal polarization, respectively. As it can be seen, if the interference received by cross-polar transmissions is small, one can recover the information data by separately equalizing each polarization. Now, in order to recover the superimposed data symbols, we compare the g-th element of both $\hat{\mathbf{s}}_{k,g,u}^{v}$ and $\hat{\mathbf{s}}_{k,g,u}^{h}$, and it is chosen the one that provides the maximum effective channel gain. More specifically, the message for the u-th user at the g-th group in the k-th cluster is retrieved according to

$$\hat{s}_{k,g,u} = [\hat{\mathbf{s}}_{k,g,u}^{p}]_g = \sum_{n=1}^{U} \alpha_{k,g,n} s_{k,g,n} + [\bar{\mathbf{H}}_{k,g,u}^{\dagger p} \mathbf{n}_{k,g,u}^{p}]_g, \tag{28}$$

where p denotes the polarization that provides the maximum effective channel gain, and $\mathbf{n}_{k,g,u}^{p} \in \mathbb{C}^{\frac{N}{2}\times 1}$ is the noise vector corresponding to the polarization p, with $p \in \{v,h\}$.

Note that such a strategy can be seen as a simple selection diversity technique that chooses the polarization that delivers the best channel condition. In this case, the offered polarization diversity can be truly explored, which implies that significant performance improvements can be obtained over the first scheme while maintaining a very low implementation complexity. The only drawback of this approach is that the maximum number of groups that can be

served in parallel is reduced by half. Finally, it is important to highlight that this strategy can also be applied to scenarios with high cross-polar interference, although one can not equalize separately each polarization, as it was done in (27), and this arises as a potential future work.

3.4 Performance Analysis

In this section, the performance of the proposed multi-polarized massive MIMO-NOMA system is evaluated. Closed-form expressions for the respective outage probability are derived, based on which an asymptotic analysis at high SNR regime is carried out. Finally, the ergodic sum-rate capacity of both approaches is also investigated.

3.4.1 SINR Analysis for the Two Proposed Approaches

After the superimposed data symbol is recovered from either Approach I or Approach II, the user employs SIC in order to retrieve its message. In order to successfully complete the SIC process, it is crucial to have knowledge about the ordering of users' effective gains. Therefore, it is assumed that the BS has complete access to the user ordering information. Moreover, it is considered that the effective channel gains are sorted out in increasing order of magnitude, meaning that the 1st user has the weakest gain and the U-th user has the strongest one. Under these assumptions, the SINRs for the two approaches are defined by the following lemmas.

Lemma 3.1. (SINR for Approach I) *During the SIC process, the SINR of the data symbol intended for the i-th weaker user that is observed at the current u-th user in the g-th group of the k-th cluster is given by*

$$\gamma^{i}_{k,g,u} = \frac{\zeta_{k,g,u}\alpha^{2}_{k,g,i}}{\zeta_{k,g,u}\mathscr{I}_{i} + \frac{1}{\rho}}, \qquad for \quad 1 \leq i \leq u \leq U. \tag{29}$$

where $\rho = \frac{1}{\sigma_n^2}$ *denotes the transmit SNR,* $\zeta_{k,g,u} = \frac{1}{\|[\tilde{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2}$ *is the effective channel gain, and* $\mathscr{I}_i$ *corresponds to the power of interfering messages, being defined by*

$$\mathscr{I}_{i} = \begin{cases} \sum_{j=i+1}^{U} \alpha^{2}_{k,g,j}, & for \quad 1 \leq i \leq u < U, \\ 0, & for \quad i = u = U, \end{cases} \tag{30}$$

Proof. Please, see Appendix A. □

Lemma 3.2. (SINR for Approach II) *Similar to the first lemma, the SINR achieved at the current u-th user while decoding the message intended to the i-th weaker user can be defined as*

$$\gamma^{i}_{k,g,u} = \frac{\zeta^{*}_{k,g,u}\alpha^{2}_{k,g,i}}{\zeta^{*}_{k,g,u}\mathscr{I}_{i} + \frac{1}{\rho}}, \qquad for \quad 1 \leq i \leq u \leq U, \tag{31}$$

where the effective channel gain is given by $\zeta^{*}_{k,g,u} = \max\{\zeta^{v}_{k,g,u}, \zeta^{h}_{k,g,u}\}$, *in which* $\zeta^{v}_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger v}_{k,g,u}]_{g,*}\|^{2}}$ *and* $\zeta^{h}_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger h}_{k,g,u}]_{g,*}\|^{2}}$. *The transmit SNR* ρ *and the power of interfering messages* $\mathscr{I}_{i}$ *are defined identically as in Lemma 3.1.*

Proof. Please, see Appendix B. □

One can see that the interference factor defined in (30) reaches its maximum value when the first user, the weakest one, decodes its message. As discussed in Subsection 2.1.3, this is expected since the first user does not decode messages from others, only its own, suffering interference from all the other users. On the other hand, assuming perfect SIC, the strongest user recovers its own message with zero interference. In addition, one can realize that the SINR expressions for Approaches I and II have a similar form, differing only in terms of the effective channel gains.

3.4.2 Outage Probability

The outage probability for the u-th user in the g-th group of the k-th cluster, denoted by $P^{\text{out}}_{k,g,u}$, can be defined as the probability of the message intended to the i-th user being received at the u-th user with a data rate less than the required target rate $\mathscr{R}_{k,g,i}$, $\forall\ 1 \leq i \leq u$, i.e.,

$$P^{\text{out}}_{k,g,u} = P[\log_{2}(1+\gamma^{i}_{k,g,u}) < \mathscr{R}_{k,g,i}], \qquad \text{for all} \quad 1 \leq i \leq u. \tag{32}$$

From (32), next, closed-form expressions for the outage probability are derived for dual-polarized massive MIMO-NOMA systems assuming the previously discussed approaches.

Proposition 3.1. (Outage Probability for Approach I) *Assuming that the users are sorted out in increasing order based on their effective channel gains, the outage probability can be derived as*

$$P^{out}_{k,g,u} = \mathscr{U}_{u} \sum_{n=0}^{U-u} (-1)^{n} \binom{U-u}{n} \frac{1}{u+n} \left[\frac{\gamma\big(\eta, (\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}\big)}{\Gamma(\eta)} \right]^{u+n}, \tag{33}$$

where $\mathscr{U}_{u} = U\binom{U-1}{u-1}$, $\eta = \frac{N}{2} - \frac{\bar{M}}{2} + 1$, $\beta_{k,g} = [\mathbf{I}_{2} \otimes (\tilde{\mathbf{B}}^{H}_{k}\tilde{\mathbf{R}}_{k}\tilde{\mathbf{B}}_{k})^{-1}]_{g,g}$ *and* $\Upsilon_{k,g,u} = \max\limits_{i\in[1,u]} \left\{ \frac{2^{\mathscr{R}_{k,g,i}}-1}{\alpha^{2}_{k,g,i} - \mathscr{I}_{i}(2^{\mathscr{R}_{k,g,i}}-1)} \right\}$.

Proof. Please, see Appendix C. □

Proposition 3.2. (Outage Probability for Approach II) *Suppose that the system experiences a small inverse XPD and assuming that the users are sorted out in increasing order based on their effective channel gains, the outage probability can be derived as*

$$P_{k,g,u}^{out} = \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{1}{u+n} \left[\frac{\gamma\left(\eta, (\chi+1)\tilde{\beta}_{k,g}\rho^{-1}\Upsilon_{k,g,u}\right)}{\Gamma(\eta)} \right]^{2(u+n)}, \tag{34}$$

where $\mathscr{U}_u$, η and $\Upsilon_{k,g,u}$ are defined identically as in Proposition 3.1, and $\tilde{\beta}_{k,g} = [(\tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k)^{-1}]_{g,g}$.

Proof. Please, see Appendix D. □

From above, note that the greater exponent in (34) influences the outage probability to decrease faster than in (33), indicating a superior performance of the Approach II.

3.4.3 Asymptotic Analysis

Even though we have derived exact expressions for the outage probability in Propositions 3.1 and 3.2, it is still complex to analyze some important performance features. Therefore, simpler expressions are desirable to investigate further the behavior of the proposed approaches. In view of this, asymptotic studies are carried out in Propositions 3.3 and 3.4, in which we derive simpler outage probability expressions and determine the diversity orders for the two approaches, as follows.

Proposition 3.3. (Asymptotic Outage Probabilities)
Asymptotic Outage Probability for Approach I: *When the transmit SNR goes to infinity, i.e., $\rho \to \infty$, the outage probability expression in (33) can be approximated by*

$$P_{k,g,u}^{out} \approx \frac{1}{\rho^{\eta u}} \frac{\mathscr{U}_u}{u} \frac{[(\chi+1)\beta_{k,g}\Upsilon_{k,g,u}]^{\eta u}}{(\eta!)^u}, \tag{35}$$

which yields a diversity order at the u-th user of

$$\mathscr{D}_u = \left(\frac{N}{2} - \frac{\bar{M}}{2} + 1\right) u. \tag{36}$$

Asymptotic Outage Probability for Approach II: *Now, when the transmit SNR goes to infinity, i.e., $\rho \to \infty$, the outage probability in (34) can be approximated by*

$$P_{k,g,u}^{out} \approx \frac{1}{\rho^{2\eta u}} \frac{\mathscr{U}_u}{u} \frac{[(\chi+1)\beta_{k,g}\Upsilon_{k,g,u}]^{2\eta u}}{(\eta!)^{2u}}, \tag{37}$$

and the diversity order of the u-th user can be expressed as

$$\mathscr{D}_u = (N - \bar{M} + 2)\, u. \tag{38}$$

Proof. Please, see Appendix E. □

From the asymptotic results, one can see that the system performance increases with the order of the users for both approaches. This behavior is in fact expected, since the higher is the user order, the better is its channel condition. Besides, it can be observed that the diversity order of the Approach I is half of the Approach II, which means that the former does not reach the same performance level of the latter one, regardless of the number of users or transmit/receive antennas.

3.4.4 Ergodic Sum-Rate

The ergodic sum-rate informs the maximum transmission sum capacity that can be achieved by a communication system. Next, we analyze the ergodic sum-rate achieved within one group, where it is assumed that all the weaker messages are successfully decoded at each user. Thus, the capacity for the u-th user can be defined based only on its own SINR, that is, based on $\gamma^{u}_{k,g,u}$. Therefore, the sum-rate for the g-th group in the k-th cluster can be defined as

$$C_{k,g} = \sum_{u=1}^{U} \log_2(1 + \gamma^{u}_{k,g,u}). \tag{39}$$

Proposition 3.4. (Ergodic Sum-Rate for Approach I) *Assuming that, during the SIC process, the messages intended to the i-th user that arrives at the u-th current user, $\forall\, i \in \{1,2,\cdots,u\}$, are correctly decoded, the ergodic sum-rate achieved by the g-th group in the k-cluster can be derived as*

$$\begin{aligned}\bar{C}_{k,g} = &\sum_{u=1}^{U} \mathscr{U}_u \sum_{n=0}^{U-u} {}^{(-1)^n}\binom{U-u}{n} \frac{(\chi+1)^{\eta}\beta_{k,g}^{\eta}}{\Gamma(\eta)^{u+n}} \\ &\times \int_0^{\infty} \log_2\left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}}\right) x^{\eta-1} e^{-(\chi+1)\beta_{k,g}x} \gamma\big(\eta, (\chi+1)\beta_{k,g}x\big)^{u-1+n} dx,\end{aligned} \tag{40}$$

where $\varepsilon_{1,u} = \rho(\alpha^2_{k,g,u} + \mathscr{I}_u)$ and $\varepsilon_{2,u} = \rho\,\mathscr{I}_u$.

Proof. Please, see Appendix F. □

Proposition 3.5. (Ergodic Sum-Rate for Approach II) *Similarly, assuming that all messages are correctly decoded during SIC, the ergodic sum-rate achieved by the g-th group in the k-*

cluster can be expressed as

$$\bar{C}_{k,g} = \sum_{u=1}^{U} \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta} \beta_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} \times \int_0^{\infty} \log_2\left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}}\right) x^{\eta-1} e^{-(\chi+1)\beta_{k,g}x} \gamma\big(\eta,(\chi+1)\beta_{k,g}x\big)^{1+2(u-1+n)} dx, \quad (41)$$

where $\varepsilon_{1,u}$ and $\varepsilon_{2,u}$ are defined exactly as in Proposition 3.4.

Proof. Please, see Appendix G. □

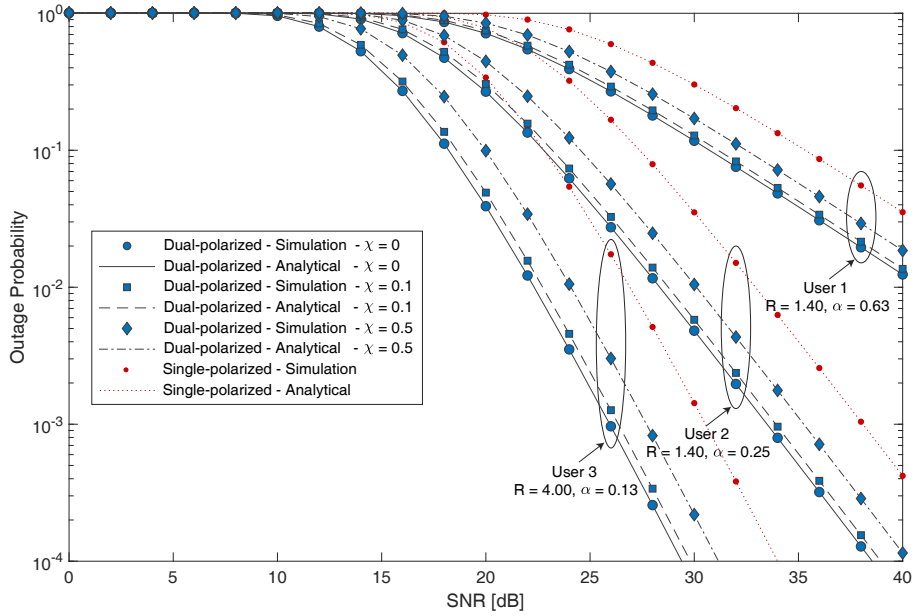
3.4.5 OMA Design

For comparison purposes, the conventional massive MIMO-OMA system is now discussed. We also consider the existence of K spatial clusters subdivided into G groups. The precoders are constructed identically as in the MIMO-NOMA system, in which the users of each group recover their respective transmitted data symbols by employing a zero forcing equalizer. However, inside each group, instead of adopting NOMA, time division multiple access is employed, meaning that each user within the group is served in a separate time slot. Consequently, the number of users that are served in parallel within one time slot in a given cluster is equal to the number of groups, i.e. G users, one from each group, are served simultaneously. Therefore, in order to provide a fair performance comparison, the number of groups in OMA is set to the same number of users within one group in the NOMA system. With this consideration, the performance of U users in NOMA will be compared with $G = U$ OMA users. Besides, the total transmission power available for the group is entirely allocated to the user in the current time slot, i.e. the user's power coefficient is set to 1.

3.5 Numerical and Simulation Results

In this section, the performance of the proposed dual-polarized massive MIMO-NOMA system is investigated. Otherwise stated, the BS is equipped with $M/2 = 50$ pairs of collocated horizontally and vertically polarized antennas arranged in a uniform linear array, where each pair of elements is separated by half of the wavelength, i.e., $\lambda/2$. For comparison purposes, single-polarized systems are equipped with the same number M of transmit antennas, with elements being also separated by $\lambda/2$. Without loss of generality, we consider the existence of $K = 4$ spatial clusters, each one containing $L = 6$ users that are subdivided into $G = 2$ groups of $U = 3$ users each, and we adjust the BS azimuth inclination angle to the direction of the cluster of interest, providing maximum array gain. The azimuth angle of the kth cluster is defined as $\theta_k = \frac{\pi}{45} + (\pi - \frac{2\pi}{45})\frac{k-1}{K-1}$, for $k = 1, \cdots, K$, where each cluster has an angular spread of 15°. Under the considered geometry, the dominant eigenvalues parameter for the channel correlation matrices of each polarization is adjusted to $r^{\star} = 16$. This makes the full matrix in (2) to have $2r^{\star} = 32$ effective eigenvalues, which is the same as that one configured to

Figure 9: Outage probability versus transmit SNR for dual-polarized massive MIMO-NOMA system under Approach 1 ($N = 4$).

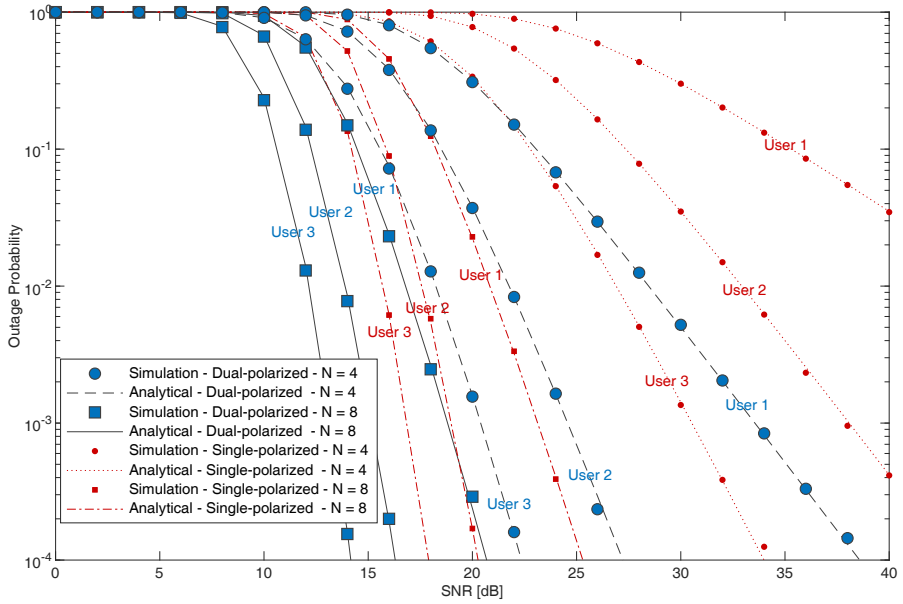


Source: Created by the author.

the single-polarized system. The power allocation coefficients of Users 1, 2 and 3 are set to $\alpha_1^2 = 0.63, \alpha_2^2 = 0.25, \alpha_3^2 = 0.12$, and the target rates to $\mathscr{R}_1 = \mathscr{R}_2 = 1.4, \mathscr{R}_3 = 4$ bits per channel use (BPCU), respectively. In addition, the number of effective data streams is set to $\bar{M} = 4$. All simulation results are generated through 2×10^5 Monte Carlo iterations, and a perfect agreement with the analytical ones is observed in all the plots.

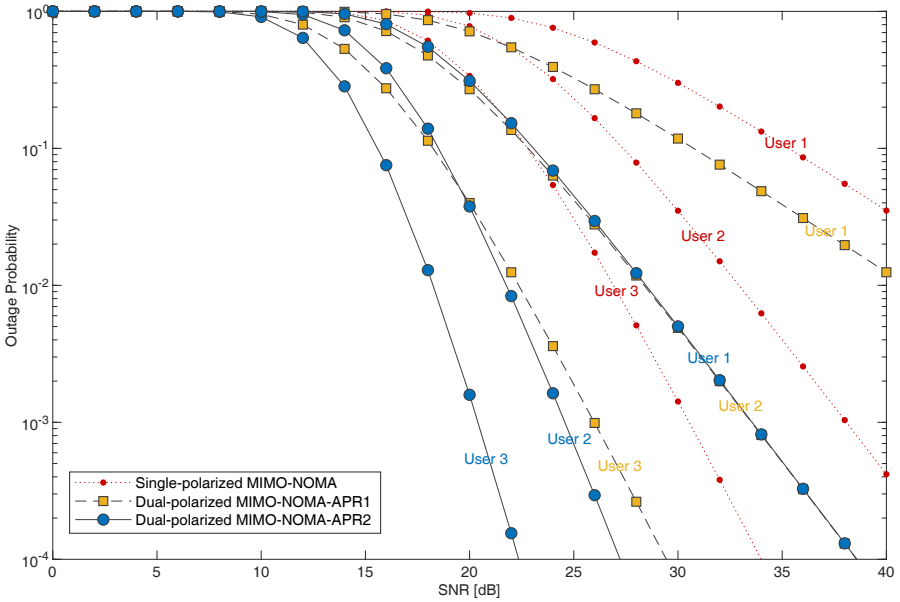
Figure 9 plots the outage probability versus transmit SNR for Approach I, in which the users are equipped with $N/2 = 2$ pairs of dual-polarized antennas. As can be realized, the system performance gets worse as χ gets higher. However, it is interesting to note that, even with a high inverse XPD value, the dual-polarized system shows superior performance than the single-polarized one, such that, for $\chi = 0.5$, the dual-polarized user with better channel conditions requires approximately 3dB less than for the single case to achieve the same level of outage. In the same way, Figure 10 shows the outage performance but now considering Approach II with a low cross-polar interference, i.e., $\chi = 0.001$. One can see the great performance improvements that this second approach can offer, confirming the insights highlighted in Section 4.3.2. For instance, when adopting 8 receive antennas, the user with strongest channel gain in the dual-polarized system can reach the same performance level of that obtained by the single-polarized user, but saving around 4dB of SNR. This advantage becomes even greater when employing 4 receive antennas, where a gain of approximately 12dB can be reached. Figure 11 compares the outage performance of the two approaches with that of the single-polarized MIMO-NOMA by setting $N = 4$ and $\chi = 0.001$. It is shown that Approach 2 has the best performance, which confirms the potential benefits that one can obtain by exploring diversity through the additional polarization dimension offered by the multi-polarized design.

Figure 10: Outage probability versus transmit SNR for dual-polarized massive MIMO-NOMA system under Approach 2 ($\chi = 0.001$).



Source: Created by the author.

Figure 11: Outage probability versus transmit SNR for single-polarized and dual-polarized MIMO-NOMA systems under Approaches 1 and 2 ($N = 4$ and $\chi = 0.001$).

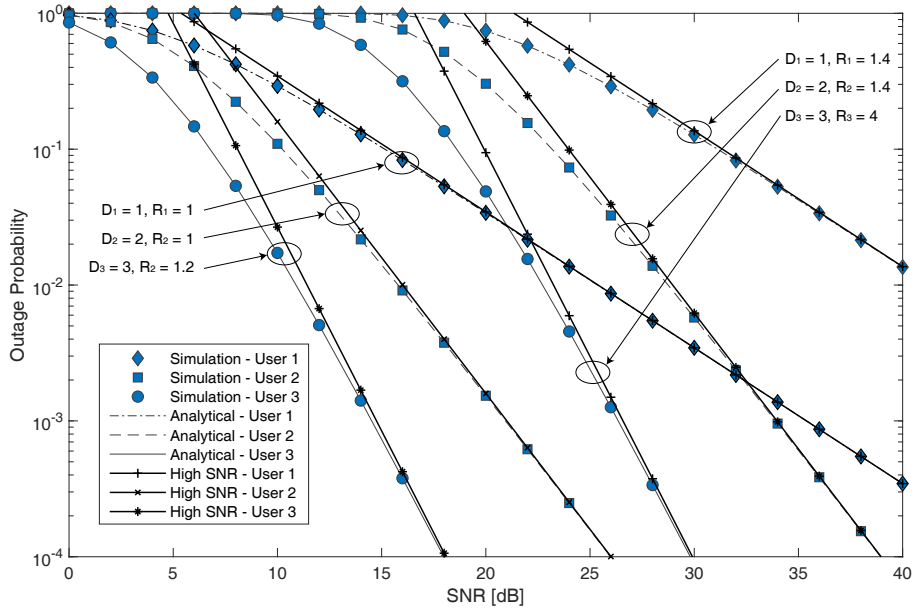


Source: Created by the author.

Also, one can observe that the gain achieved by Approach 1 over the single-polarized case can go up to 4 dB.

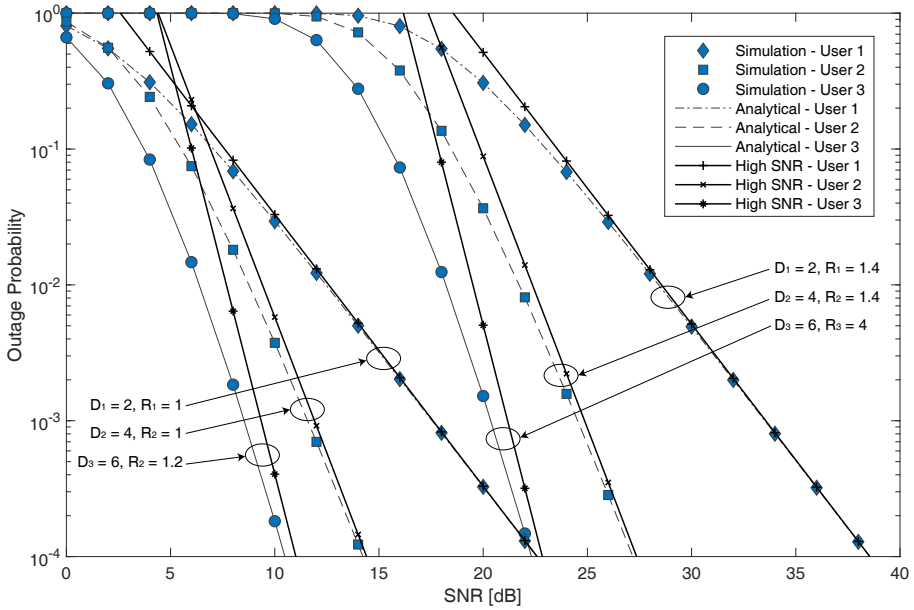
Figures 12 and 13 show the exact outage probability curves along with their respective asymptotic behaviors for Approaches 1 and 2, respectively, in which a perfect agreement at high SNR regions is observed among the curves. One can observe that the diversity order for both approaches increases with the user order. Besides, changing the values of target rates does

Figure 12: Exact and asymptotic outage probability curves for dual-polarized massive MIMO-NOMA system under Approach 1 ($N = 4$ and $\chi = 0.1$).



Source: Created by the author.

Figure 13: Exact and asymptotic outage probability curves for dual-polarized massive MIMO-NOMA system under Approach 2 ($N = 4$ and $\chi = 0.001$).

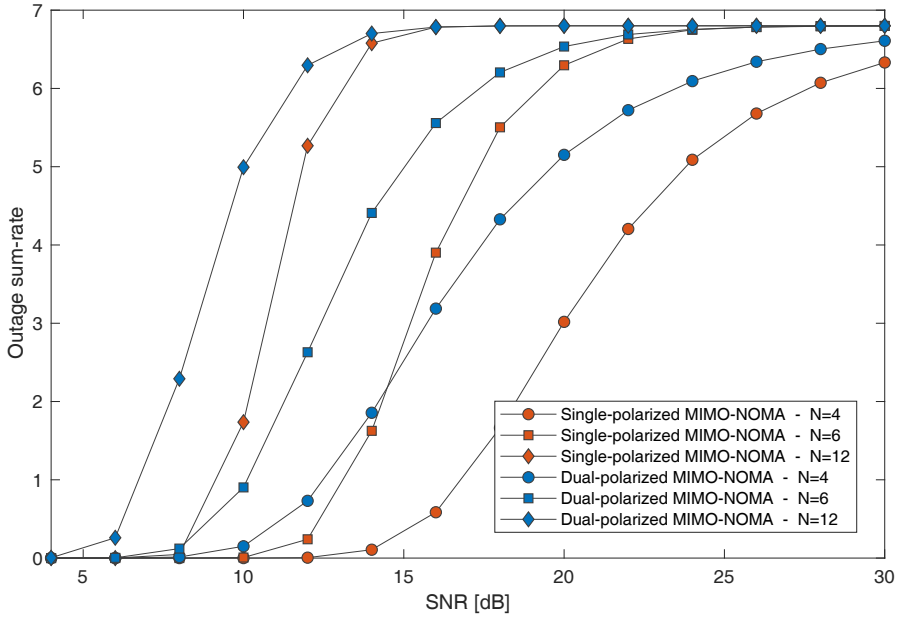


Source: Created by the author.

not affect the diversity orders. These behaviors are in complete agreement with the analytical derivation of Proposition 3.3. In addition, as anticipated in Section 3.4.3, the dual-polarized system under Approach 2 exhibits higher diversity orders than the obtained with approach 1.

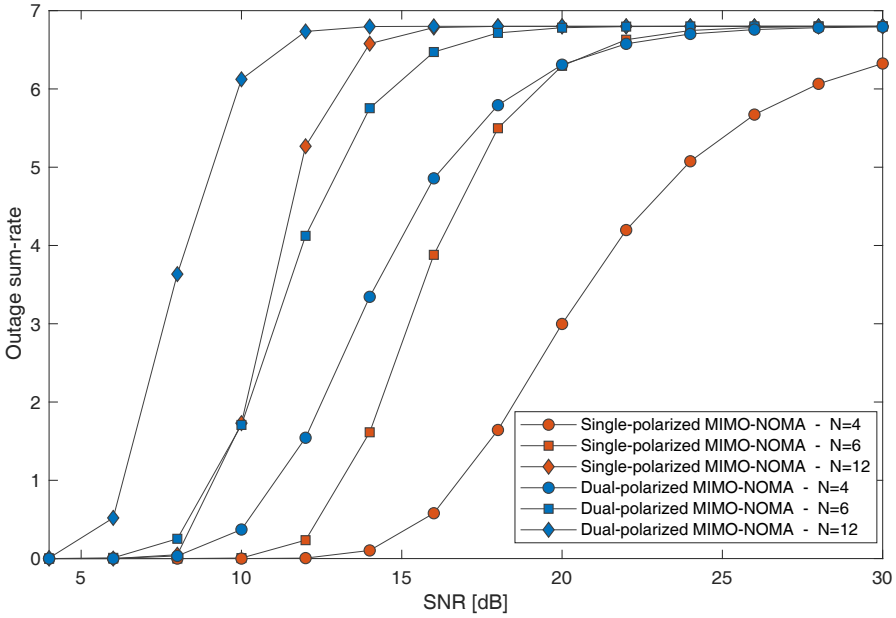
In Figures 14 and 15, the outage sum-rate, obtained as $\sum_{n=1}^{U}(1 - P_{k,g,n}^{\text{out}})\mathscr{R}_{k,g,n}$, is investigated considering different numbers of receive antennas. This metric computes the sum of the users' throughput achieved when the BS is transmitting at fixed target rates. As can be

Figure 14: Outage sum-rate for single-polarized and dual-polarized MIMO-NOMA system under Approach 1 ($\chi = 0.1$).



Source: Created by the author.

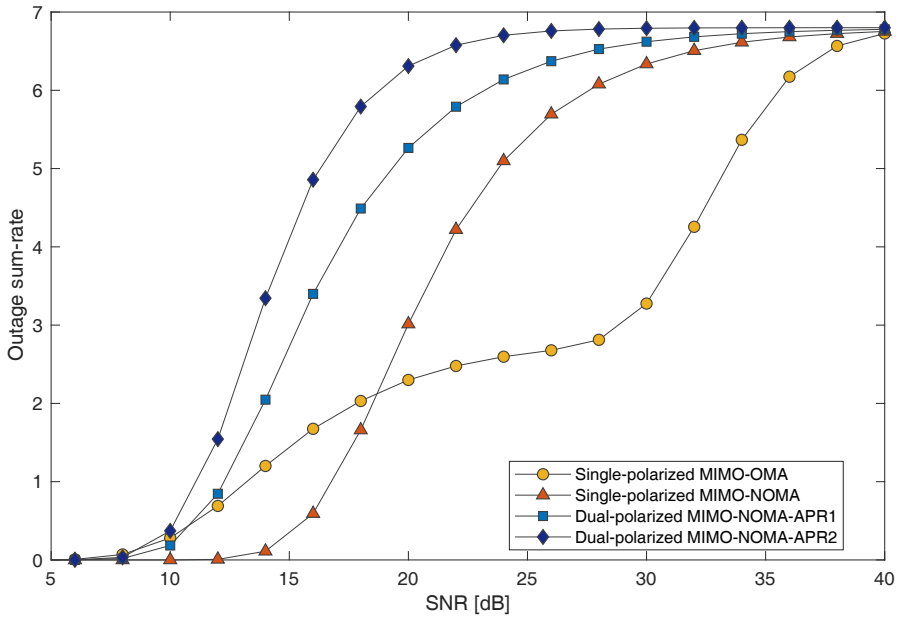
Figure 15: Outage sum-rate curves for single-polarized and dual-polarized MIMO-NOMA system under Approach 2 ($\chi = 0.001$).



Source: Created by the author.

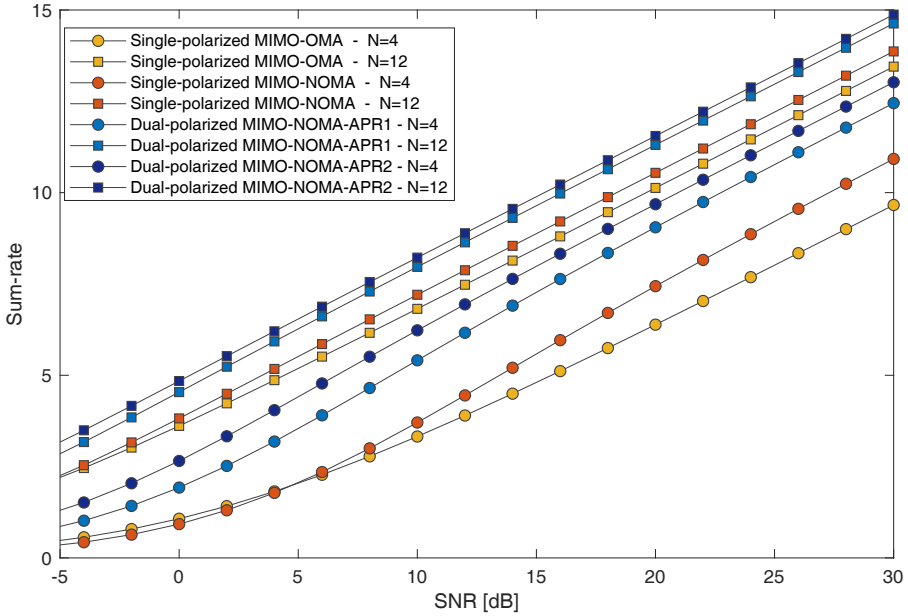
observed, antenna polarization can provide significant spectral improvements, which the gains attained with Approach 2 being even more impressive. For instance, in Figure 15, for $N = 4$ and a SNR of 16dB, the dual-polarized scheme can achieve a throughput of 4.8 BPCU, what is about 8 times greater than that of the single-polarized system with users equipped with the same number of receive antennas, and 1 BPCU higher than what is obtained with $N = 6$. Despite the

Figure 16: Outage sum-rate curves for single-polarized MIMO-OMA/MIMO-NOMA systems and dual-polarized MIMO-NOMA systems under Approaches 1 and 2 ($N = 4$ and $\chi = 0.001$).



Source: Created by the author.

Figure 17: Simulated ergodic sum-rate for various system setups ($\chi = 0.001$).



Source: Created by the author.

gains, one can realize that with the increase of the number of receive antennas, the SNR needed for both systems to exhibit identical performances is decreased. This can be justified through the systems' diversity order expressions. Figure 16 compares the outage sum-rate performance of the dual-polarized MIMO-NOMA system, under both proposed approaches, with the single-polarized OMA and NOMA cases. One can observe that Approach 2 outperforms Approach 1, in which the obtained throughput gap can reach up to 1.5 BPCU for a SNR value of 16dB. In

addition, the dual-polarized NOMA systems outperform the classic OMA scheme in almost all considered SNR range, while the single-polarized MIMO-NOMA case becomes better only for SNR values higher than 19dB, which demonstrates again the benefits of polarization.

Finally, Figure 17 investigates the ergodic sum-rate for the single-polarized OMA and NOMA schemes as well as for the proposed dual-polarized NOMA design with Approaches 1 and 2. One can observe the remarkable improvements that the dual-polarized MIMO-NOMA system can provide, in which the dual-polarized scheme outperforms all the single-polarized systems.

3.6 Chapter Summary

In this chapter, we have investigated the application of dual-polarized antenna arrays in massive MIMO-NOMA systems under highly correlated multi-cluster scenarios. Two precoder designs were proposed and investigated, and a detailed analytical analysis in terms of outage probability and outage/ergodic sum-rate was carried out. An asymptotic analysis in high SNR regime was also performed, in which the diversity gains for the two proposed approaches were determined. In addition, numerical and simulation results were provided to validate the analytical framework and to demonstrate the advantages of the proposed dual-polarized designs.

4 MASSIVE MIMO-NOMA NETWORKS WITH SUCCESSIVE SUB-ARRAY ACTIVATION

This chapter presents the second part of the original contributions developed in this dissertation, in which a novel diversity strategy for massive MIMO-NOMA systems is proposed. The chapter is organized as follows. In Section 4.2, the massive MIMO-NOMA system model with multiple antenna sub-arrays at the BS is introduced. A detailed description of the proposed diversity scheme protocol and the design of the beamforming and detection matrices are also presented. In Section 4.3, the analytical analysis of the system performance is carried out. Details about conventional MIMO-OMA and MIMO-NOMA schemes, used for performance comparison, are described in Section 4.4. Numerical and simulation results along with comprehensive discussions are presented in Section 4.5, while the final considerations are drawn in Section 4.6.

4.1 Motivation and Contributions

The exploration of all forms of diversity, in frequency, code or time domains, will be essential for satisfying the high quality of service requirements of 5G networks. However, there is a lack of related contributions, and only a very limited number of works investigates diversity techniques in massive MIMO-NOMA systems. This motivates further studies in this field of research. Owing to this fact, by combining concepts of time and space diversity, we propose a novel low-complexity scheme with the potential of improving the outage performance of each user in a massive MIMO-NOMA deployment. A single-cell multi-cluster downlink scenario is assumed, where the base-station sends redundant symbols through multiple sub-arrays to multi-antenna receivers (only single-polarized antennas are considered). More details and main contributions presented in this chapter are summarized as follows:

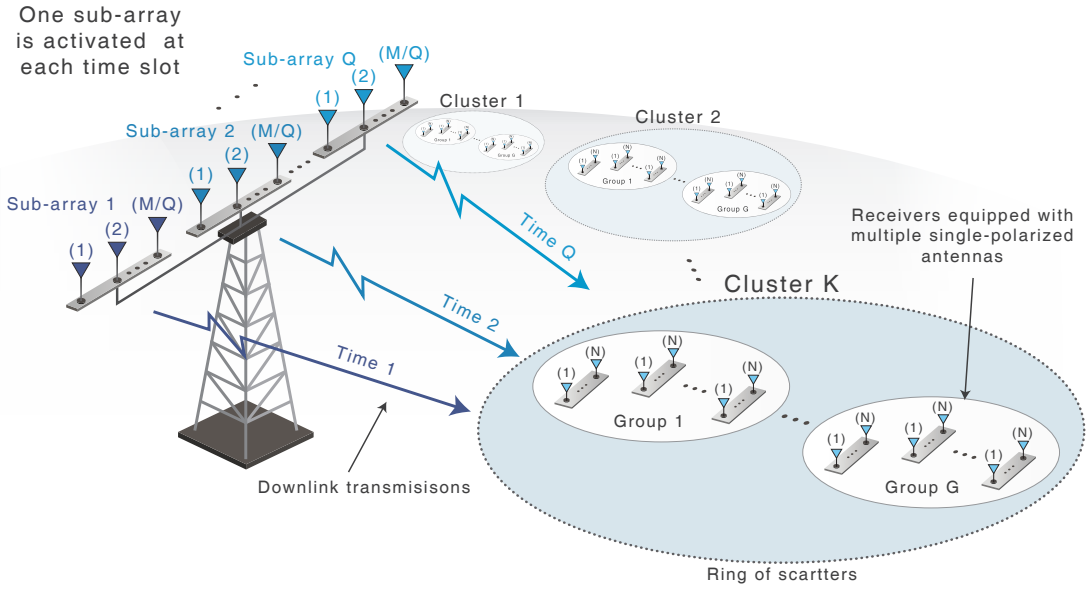
- We design a novel open-loop diversity technique that enables each of the users to improve their reception performance without the need for transmitting back their fast varying channel matrices. This is achieved by successively activating antenna sub-arrays at the BS and exploring space diversity in different instants of time at the users' side. With this strategy, the system latency can still be maintained low.
- We adopt two-stage precoders for each of the antenna sub-arrays. The outer precoding matrices are constructed based only on the channel covariance matrices, and the inner precoders are only intended for assigning the messages for each of the NOMA groups. These choices also do not require knowledge of the fast varying channels. Therefore, in our proposal, either for providing diversity or for constructing the precoders, only the slowly varying covariance matrices are needed, which is attractive for applications where feedback is expensive or very limited.
- An in-depth analytical analysis in the proposed design is performed, in which an exact closed-form expression for the outage probability is derived. The system behavior at high

SNR regime is also investigated, where an asymptotic outage approximation is attained, enabling us to determine the system diversity order. Furthermore, we investigate and obtain a closed-form expression for the ergodic sum-rate considering a particular case.

- Numerical and simulation results are presented to corroborate the theoretical development, and insightful discussions are provided. In particular, our results show that the proposed strategy outperforms conventional full array massive MIMO-NOMA and MIMO-OMA systems operating in time diversity mode in all considered scenarios.

4.2 System Model and Protocol Description

Figure 18: System model. The BS is equipped with multiple antennas subdivided into multiple sub-arrays.



Source: Created by the author.

Consider a scenario where a single BS equipped with a linear array of M antennas is transmitting to multiple users. Each user employs N receive antennas, with $M \gg N$. Besides, the users are assumed to be confined within K rich scattering clusters, following the one ring geometrical model of Subsection 2.2.2. In each spatial cluster, there are G groups, each one containing U users that are multiplexed through power-domain NOMA. It is also assumed that the users require a reliable data reception. To fulfill this requirement, we propose a novel diversity strategy by exploring space and time dimensions. Firstly, at the BS, the transmit antennas are equally divided into Q sub-arrays, i.e., we create Q partitions of M/Q antenna elements, as shown in Figure 18. Due to this structure, M must be a multiple of Q. In addition, we adjust the separation distance between two adjacent sub-arrays to be greater than half of the wavelength, i.e., greater than $\lambda/2$, so that the channels among sub-arrays become uncorrelated. Within each sub-array, we set the inter-antenna space separation to be exactly $\lambda/2$ and we consider correlation between antenna elements. Then, in order to improve reliability, we

configure the system to send Q replicas of the same symbol. More specifically, each symbol replica is transmitted by sequentially activating each sub-array. Consequently, the transmission of one symbol is performed within Q instants of time. The proposed strategy will be called as successive sub-array activation (SSAA). Note that, since the sub-arrays are uncorrelated, each transmission will propagate through different paths, regardless of the separation time between two retransmissions. Therefore, diversity is achieved through the space dimension.

As one can realize, differently from conventional time diversity schemes, that need to wait for a whole coherent time to retransmit the data, our proposed system can operate with very fast transmission rates. The time needed to retransmit the symbol replicas must be just enough to the receiver distinguishes the signals from each sub-array. As a result, a very low latency can still be achieved, while guaranteeing an enhanced performance. Our diversity strategy is also energy efficient, since, regardless of the number of transmitted replicas, the total number of antennas activated during all retransmissions remains constant, i.e., a total of M antennas is used to transmit Q symbol replicas. In addition to these advantages, since only one sub-array is activated at a time, it is possible to reduce the number of dedicated electronic components that are connected to the antenna elements, known as radio frequency (RF) chains. More specifically, by using simple switches [81], the number of dedicated RF chains could be reduced to M/Q, i.e., the same number of antennas in a sub-array, without degrading the system performance, what would lead to a decrease in hardware cost and to lower power consumption.

4.2.1 Channel Model

We consider that all users within the k-th spatial cluster share the same channel covariance matrix $\bar{\mathbf{R}}_k = \mathbf{I}_Q \otimes \mathbf{R}_k \in \mathbb{C}^{M\times M}$, where $\mathbf{R}_k \in \mathbb{C}^{\frac{M}{Q}\times\frac{M}{Q}}$ corresponds to the covariance matrix of each sub-array. Note that, for simplicity, we assume identical covariance matrices among sub-arrays. Considering the proposed design, the BS transmits each symbol in Q instants of time, where each replica propagates independently and experiences different fast fading channels. Besides, it is assumed a perfect downlink channel estimation at the users' side. Under these considerations, each user will acquire Q distinct channel matrices, each one corresponding to a different transmission. For mathematical convenience, the channel matrices belonging to the u-th user in the g-th group of the k-th cluster are organized in the following block diagonal arrangement

$$\bar{\mathbf{H}}_{k,g,u} = \begin{bmatrix} \mathbf{H}^1_{k,g,u} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{H}^Q_{k,g,u} \end{bmatrix} \in \mathbb{C}^{M\times QN}, \tag{42}$$

where $\mathbf{H}^q_{k,g,u} \in \mathbb{C}^{\frac{M}{Q}\times N}$ represents the channel matrix obtained during reception of the message transmitted by the q-th antenna sub-array. By invoking the Karhunen-Loeve transformation, we

can decompose the sub-matrices of $\bar{\mathbf{H}}_{k,g,u}$ as follows

$$\bar{\mathbf{H}}_{k,g,u} = \begin{bmatrix} \mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,g,u}^{1} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,g,u}^{Q} \end{bmatrix}, \tag{43}$$

where $\boldsymbol{\Lambda}_k$ stands for the diagonal matrix of dimension $r_k^\star \times r_k^\star$ composed by the $r_k^\star$ decreasing nonzero eigenvalues of $\mathbf{R}_k$, $\mathbf{U}_k \in \mathbb{C}^{\frac{M}{Q} \times r_k^\star}$ represents the matrix formed by eigenvectors of $\mathbf{R}_k$, and $\mathbf{G}_{k,g,u}^{q} \in \mathbb{C}^{r_k^\star \times N}$ is the fast varying channel matrix corresponding to the q-th sub-array. Considering a NLOS communication, the entries of $\mathbf{G}_{k,g,u}^{q}$ are modeled as i.i.d. complex Gaussian distributed random variables with zero mean and unit variance.

Then, after the BS superposes the messages of all users within each group and transmit the Q successive replicas, over each sub-array, through the fast fading channels, the users observe the following signal

$$\mathbf{y} = \sum_{n=1}^{K} \bar{\mathbf{H}}_{n,g,u}^{H} \bar{\mathbf{B}}_n \sum_{i=1}^{G} \bar{\mathbf{p}}_{n,i} \sum_{j=1}^{U} \alpha_{n,i,j} s_{n,i,j} \in \mathbb{C}^{QN \times 1}, \tag{44}$$

where $\bar{\mathbf{B}}_n \in \mathbb{C}^{M \times Q\bar{M}}$ is the preprocessing matrix designed to remove inter-cluster interference, with $\bar{M}$ being the parameter that defines the virtual channel dimension, $\bar{\mathbf{p}}_{n,i} \in \mathbb{C}^{Q\bar{M} \times 1}$ is the precoding vector responsible for assigning the superposed messages to its respective groups, $\alpha_{n,i,j}$ is the power allocation coefficient, and $s_{n,i,j}$ is the message intended for the user j in the i-th group of the n-th cluster.

4.2.2 Precoder Design

In this subsection, we present details for the construction of the precoding matrices. Similarly as in the dual-polarized setups of Chapter 3, in this chapter, we choose to project the precoders not to depend on the fast varying channel matrices. Consequently, the BS will only need to estimate the covariance matrices of the lower dimension antenna sub-arrays, which provides an even further reduction in the overall feedback overhead.

Considering the proposed antenna structure, the outer precoder $\bar{\mathbf{B}}_k$, belonging to the k-th cluster, can be arranged in the following block diagonal matrix

$$\bar{\mathbf{B}}_k = \begin{bmatrix} \mathbf{B}_k & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_k \end{bmatrix}, \tag{45}$$

where $\mathbf{B}_k \in \mathbb{C}^{\frac{M}{Q} \times \bar{M}}$ denotes the precoding sub-matrix that is designed based on the slowly varying channel covariance matrix of each sub-array.

The sub-precoders $\mathbf{B}_k$ can be constructed via similar procedures as the carried out in Section 3.3. Therefore, by also assuming that $r_1^\star = \ldots = r_K^\star = r^\star$, $r_1 = \cdots = r_K = r$, and concatenating the left eigenvectors of interfering clusters to form the matrix $\mathbf{U}_k^- = [\mathbf{U}_1, \cdots, \mathbf{U}_{k-1}, \mathbf{U}_{k+1}, \cdots, \mathbf{U}_K] \in \mathbb{C}^{\frac{M}{Q} \times (K-1)r^\star}$, we explore the null space spanned by $\mathbf{U}_k^-$ to build $\mathbf{B}_k$. More specifically, let $\mathbf{E}_k^0 \in \mathbb{C}^{\frac{M}{Q} \times \frac{M}{Q} - (K-1)r^\star}$ be the matrix that collects the last $\frac{M}{Q} - (K-1)r^\star$ columns of the left eigenvectors of $\mathbf{U}_k^-$. Then, we define the projected channel $\tilde{\mathbf{H}}_{k,g,u}^q = (\mathbf{E}_k^0)^H \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,g,u}^q$, which is orthogonal to the eigen-space spanned by interfering clusters and has covariance matrix given by $\tilde{\mathbf{R}}_k = (\mathbf{E}_k^0)^H \mathbf{R}_k \mathbf{E}_k^0$.

Let now $\mathbf{F}_k^{(1)} \in \mathbb{C}^{\frac{M}{Q} - (K-1)r^\star \times \bar{M}}$ be the first $\bar{M}$ columns of the left eigenvectors of $\tilde{\mathbf{R}}_k$. Then, finally, the outer precoding matrix for each antenna sub-array can be obtained as

$$\mathbf{B}_k = \mathbf{E}_k^0 \mathbf{F}_k^1 \in \mathbb{C}^{\frac{M}{Q} \times \bar{M}}, \tag{46}$$

in which the constraints $K \leq \bar{M} \leq \left(\frac{M}{Q} - (K-1)r^\star\right)$ and $\bar{M} \leq r^\star \leq r$ must be fulfilled. Due to these constraints, in this chapter, we configure the dominant eigenvalues as
$r^\star = \min\left\{r, \left\lfloor (\frac{M}{Q} - \bar{M}) \frac{1}{K-1} \right\rfloor\right\}$.

Now, we focus on the design of the inner precoding vector. Due to the multi-array architecture, $\bar{\mathbf{p}}_{k,g}$ is composed by Q sub-vectors, which can be represented as follows

$$\bar{\mathbf{p}}_{k,g} = [(\mathbf{p}_{k,g}^1)^T, \cdots, (\mathbf{p}_{k,g}^Q)^T]^T, \tag{47}$$

where $\mathbf{p}_{k,g}^q$ is the precoding sub-vector corresponding to the q-th sub-array, and it is responsible for assigning the data message for the g-th group in the k-th cluster. Since each sub-array contains the same number of antenna elements, we have that $\mathbf{p}_{k,g}^1 = \mathbf{p}_{k,g}^2 = \cdots = \mathbf{p}_{k,g}^Q$. Therefore, we define the sub-precoders as

$$\mathbf{p}_{k,g}^q = [\ \mathbf{0}_{1\times(g-1)}\ ,\ 1\ ,\ \mathbf{0}_{1\times(\bar{M}-g)}\ ]^T, \quad \forall q = 1, \cdots, Q. \tag{48}$$

With this choice, note that the g-th effective data stream of each sub-array is associated with the g-th group, which enables each group to receive Q copies of the same superposed symbol.

4.2.3 Signal Reception

Considering that the outer precoder, designed in the last subsection, perfectly cancels the inter-cluster interferences, and after all Q successive transmissions have been received, the data signal at the u-th user in the g-th group in the k-th cluster can be structured as

$$\mathbf{y}_{k,g,u} = \bar{\mathbf{H}}_{k,g,u}^H \bar{\mathbf{B}}_k \sum_{i=1}^{G} \bar{\mathbf{p}}_{k,i} \sum_{j=1}^{U} \alpha_{k,i,j} s_{k,i,j} + \left[\mathbf{n}_{k,g,u}^1, \cdots, \mathbf{n}_{k,g,u}^Q\right]^T, \tag{49}$$

where $\mathbf{n}_{k,g,u}^{q} \in \mathbb{C}^{N\times 1}$ is a complex Gaussian noise vector obtained during reception of the signal transmitted by the q-th sub-array, with entries having zero-mean and variance σ_n^2.

Then, in order to the separate the superposed data symbols intended for each group, the zero-forcing receiver described in Subsection 2.2.5 is employed in the users' terminals. Mathematically, the signal obtained in (49) is filtered through the following detection matrix

$$\bar{\mathbf{H}}_{k,g,u}^{\dagger} = \begin{bmatrix} \mathbf{H}_{k,g,u}^{1\dagger} & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{k,g,u}^{Q\dagger} \end{bmatrix}, \tag{50}$$

in which $\mathbf{H}_{k,g,u}^{q\dagger} = (((\mathbf{H}_{k,g,u}^{q})^H \mathbf{B}_k)^H (\mathbf{H}_{k,g,u}^{q})^H \mathbf{B}_k)^{-1} ((\mathbf{H}_{k,g,u}^{q})^H \mathbf{B}_k)^H$ is the pseudo-inverse of the virtual channel observed during the q-th reception, where we suppose that $\bar{M} \leq N$. After zero-forcing filtering, the channel distortion is completely removed and the users obtain a noise corrupted version of the signals transmitted by each antenna sub-array. With the proposed inner precoder in (48), the detected symbol vector can be represented by

$$\begin{aligned} \hat{\mathbf{s}}_{k,g,u} &= \left[\mathbf{s}_k^1, \cdots, \mathbf{s}_k^Q\right]^T + \bar{\mathbf{H}}_{k,g,u}^{\dagger} \left[\mathbf{n}_{k,g,u}^1, \cdots, \mathbf{n}_{k,g,u}^Q\right]^T \\ &= \left[\begin{bmatrix} \sum_{j=1}^{U} \alpha_{k,1,j} s_{k,1,j} \\ \vdots \\ \sum_{j=1}^{U} \alpha_{k,G,j} s_{k,G,j} \end{bmatrix}, \cdots, \begin{bmatrix} \sum_{j=1}^{U} \alpha_{k,1,j} s_{k,1,j} \\ \vdots \\ \sum_{j=1}^{U} \alpha_{k,G,j} s_{k,G,j} \end{bmatrix}\right]^T + \bar{\mathbf{H}}_{k,g,u}^{\dagger} \begin{bmatrix} \mathbf{n}_{k,g,u}^1 \\ \vdots \\ \mathbf{n}_{k,g,u}^Q \end{bmatrix}, \end{aligned} \tag{51}$$

where $\mathbf{s}_k^q$ is the vector of superposed data symbols transmitted through the q-th sub-array to users within the k-th cluster, in which $\mathbf{s}_k^1 = \cdots = \mathbf{s}_k^Q$.

Note that, the users within the g-th group can recover their data messages through the g-th element of any of the Q sub-vectors in (51), which can be accomplished by employing any diversity combining technique. In our design, due to low complexity, we simply select the symbol from the sub-vector that delivers the best effective channel gain.

4.3 Performance Analysis

In this section, the performance of the proposed massive MIMO-NOMA design operating with the SSAA scheme is investigated. First, we analyze the SINR that is experienced by the users during each NOMA decoding. Then, based on the SINR result, the system outage probability is evaluated, in which an exact closed-form expression is obtained. In addition, a high SNR asymptotic analysis is also performed, where we identify the diversity order obtained with the proposed SSAA strategy. Aiming to achieve further insights about the system behavior, the ergodic sum-rate is also studied. In particular, considering a special case, we derive an exact ergodic sum-rate expression.

4.3.1 SINR Analysis for the SSAA Diversity Scheme

Similarly as considered in Section 3.4, here, we also assume that the users are sorted out in ascending order based on the magnitude of their effective channel gains, which are supposed to be perfectly known at the BS. Under this assumption, after all the SIC decodings are completed, the u-th user in the g-th group will observe the following data symbol

$$\hat{s}_{k,g,u} = \underbrace{\alpha_{k,g,u} s_{k,g,u}}_{\text{symbol of interest}} + \underbrace{\sum_{j=u+1}^{U} \alpha_{k,g,j} s_{k,g,j}}_{\text{interference}} + \underbrace{[\mathbf{H}_{k,g,u}^{q\dagger} \mathbf{n}_{k,g,u}^{q}]_g}_{\text{noise}}. \tag{52}$$

where $q \in \{1, 2, \cdots, Q\}$ corresponds to the sub-array that achieves the maximum effective channel gain among all Q transmissions. From (52), the SINR obtained at the u-th user while recovering its message is defined in Lemma 1.

Lemma 4.1. *Supposing that the users recover its desired message from the sub-array that delivers the best effective gain, the SINR of the u-th user in the g-th group while decoding the message intended for the i-th user,* $1 \leq i \leq u \leq U$*, is given by*

$$\gamma_{k,g,u}^{i} = \frac{\bar{\varsigma}_{k,g,u} \alpha_{k,g,i}^2}{\bar{\varsigma}_{k,g,u} \mathscr{I}_i + \frac{1}{\rho}}, \qquad \textit{for} \quad 1 \leq i \leq u \leq U, \tag{53}$$

where $\bar{\varsigma}_{k,g,u} = \max\left\{\varsigma_{k,g,u}^{1}, \cdots, \varsigma_{k,g,u}^{Q}\right\}$ *denotes the effective channel gain, with* $\varsigma_{k,g,u}^{q} = \frac{1}{\|[\mathbf{H}_{k,g,u}^{q\dagger}]_{g*}\|^2}$, $1 \leq q \leq Q$. $\rho = \frac{1}{\sigma_n^2}$ *represents the transmit SNR, and* $\mathscr{I}_i$ *corresponds to the power of interfering users, which is defined by*

$$\mathscr{I}_i = \begin{cases} \sum_{j=i+1}^{U} \alpha_{k,g,j}^2, & \textit{for} \quad 1 \leq i \leq u < U, \\ 0, & \textit{for} \quad i = u = U, \end{cases} \tag{54}$$

Proof. Please, see Appendix H. □

4.3.2 Outage Probablity

If the achieved data rate at the u-th user while decoding the i-th message obtained from the sub-array that delivered the highest SINR is less than the required rate $\mathscr{R}_{k,g,i}$, this user will not be able to retrieve its message, leading to an outage event. Thus, the outage probability for the system operating with the SSAA scheme can be defined identically as in (32), i.e.,

$$P_{k,g,u}^{\text{out}} = P[\log_2(1 + \gamma_{k,g,u}^{i}) < \mathscr{R}_{k,g,i}], \quad \forall i = 1, \cdots, u. \tag{55}$$

A closed-form expression for the outage probability achieved in the massive MIMO-NOMA setup with the proposed diversity strategy is provided in the following proposition.

Proposition 4.1. *Supposing that the effective channel gains are ordered as* $\bar{\varsigma}_{k,g,1} < \bar{\varsigma}_{k,g,2} < \cdots < \bar{\varsigma}_{k,g,u}$*, the outage probability for the massive MIMO-NOMA system operating with the proposed SSAA scheme can be derived as*

$$P_{k,g,u}^{out} = \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} \frac{(-1)^j}{u+j} \left[\frac{\gamma\left(N-\bar{M}+1, \rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}\right)}{\Gamma(N-\bar{M}+1)} \right]^{Q(u+j)}, \quad (56)$$

where $\mathscr{U}_u = U\binom{U-1}{u-1}$ *and* $\Upsilon_{k,g,u} = \max\limits_{1\le i\le k} \left\{ \frac{2^{\mathscr{R}_{k,g,i}}-1}{\alpha_{k,g,i}^2 - \mathscr{I}_i(2^{\mathscr{R}_{k,g,i}}-1)} \right\}$.

Proof. Please, see Appendix I. □

From Proposition 4.1, we can obtain some insights about the performance of our proposed diversity scheme. It can be noticed that the expression in (56) is a monotonically decreasing function of the transmit SNR ρ, in which the exponent determines how fast the outage probability decreases, i.e., greater exponents lead to a faster decrease. This suggests that by increasing the number of sub-arrays Q and, consequently, increasing the exponent in (56), the system performance might be improved.

4.3.3 Asymptotic Analysis

In order to gain further insights from the derived outage expression in Proposition 4.1, an asymptotic analysis is now performed.

Proposition 4.2. *When the transmit SNR approaches infinity, (56) can be approximated by*

$$P_{k,g,u}^{out} \approx \frac{\mathscr{U}_u}{u} \frac{[\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}]^{(N-\bar{M}+1)Qu}}{\rho^{(N-\bar{M}+1)Qu}[(N-\bar{M}+1)!]^{Qu}}. \quad (57)$$

As a result, the u-th user experiences the following diversity order

$$\mathscr{D}_u = (N-\bar{M}+1)\,Qu. \quad (58)$$

Proof. Please, see Appendix J. □

From (58), we see that, for a fixed number of N and $\bar{M}$, the diversity order achieved with our proposed system design scales with the increase in the number of sub-arrays. This means that, when the SNR is high, i.e., $\rho \to \infty$, our multi-array system achieves superior performance than the conventional MIMO-NOMA counterpart, whose diversity order can be obtained in [7]. Another detail that can be observed in (58) is that the diversity order increases with the order of the user, what is indeed expected.

4.3.4 Ergodic Sum-Rate

In this subsection, we investigate the ergodic sum-rate for the proposed massive MIMO-NOMA system operating with the SSAA strategy. Here, we assume that the u-th user can successfully decode all symbols intended for the i-th weaker user, $\forall i = 1, \cdots, u$, so that, the instantaneous capacity of each user will depend only on the SINR achieved while decoding its own message. Under this assumption, the ergodic sum-rate for the g-th group, after the Q successive receptions, can be evaluated by

$$\bar{C}_{k,g} = \mathbb{E}\left[\sum_{u=1}^{U} \log_2(1+\gamma_{k,g,u}^{u})\right]. \tag{59}$$

A general solution for the expectation in (59) is presented in Proposition 4.3.

Proposition 4.3. *Under the assumption of perfect SIC decoding, the ergodic sum-rate for the g-th group achieved with the SSAA strategy can be attained by*

$$\begin{aligned}\bar{C}_{k,g} = \sum_{u=1}^{U} \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} (-1)^j \frac{Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}^{N-\bar{M}+1}}{\Gamma(N-\bar{M}+1)^{Q(u+j)}} \int_0^{\infty} \log_2\left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}}\right) \\ \times x^{N-\bar{M}} e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \gamma(N-\bar{M}+1, x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^{Q(u+j)-1} dx.\end{aligned} \tag{60}$$

where $\varepsilon_{1,u} = \rho(\alpha_{k,g,u}^2 + \mathscr{I}_u)$ *and* $\varepsilon_{2,u} = \rho\mathscr{I}_u$.

Proof. Please, see Appendix K. □

As one can notice, the ergodic sum-rate expression of Proposition 4.3 is not a closed-form solution. The reason for this is that the integral in (60) is quite challenging to solve. Even so, it is still useful, since (60) can be evaluated very efficiently through numerical methods. An exact and simpler analytical solution for (59) can be obtained if we consider the case in which the number of receive antennas is equal to the number of effective data streams, i.e., $N = \bar{M}$, as derived in Proposition 4.4.

Proposition 4.4. *For the particular case when* $N = \bar{M}$*, a closed form solution for the ergodic sum-rate can be obtained as*

$$\begin{aligned}\bar{C}_{k,g} = \sum_{u=1}^{U} \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} \frac{(-1)^j Q}{(i+1)\ln(2)} \binom{Q(u+j)-1}{i}^{(-1)^i} \\ \times \left[Ei\left(-\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{2,u}}\right) e^{\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{2,u}}} \right. \\ \left. - Ei\left(-\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{1,u}}\right) e^{\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{1,u}}} \right].\end{aligned} \tag{61}$$

where $\varepsilon_{1,u}$ *and* $\varepsilon_{2,u}$ *are defined in the same way as in Proposition 4.3.*

Proof. Please, see Appendix L. □

4.4 Schemes for Performance Comparison

In order to show the advantages of our proposed multi-array strategy, we compare its performance with some conventional schemes, with and without the exploration of time diversity. For the implementation of these schemes, we consider the same geometrical scenario as the adopted in the SSAA design, in which we assume the existence of K scattering spatial clusters with G groups of U users. In addition, for a fair comparison, the clusters are considered to be located within the same azimuth angles and have the same angular spread as in SSAA. The main difference between the conventional schemes and the system operating with the SSAA approach is that all antennas are activated at each transmission, i.e., the full transmit array is activated. Besides, for the systems that employ time diversity, in order to obtain different channel responses, the time separation between retransmissions must be greater than the channel coherent time. More details about each particular scheme are provided next.

MIMO-NOMA with time diversity: This first scheme consists of a conventional massive MIMO-NOMA system, similar as in [7], but with the difference that each superposed symbol intended for each NOMA group is redundantly transmitted over T instants of time. In this implementation, the users recover their symbols by selecting one of the T receptions, the one that achieves the highest effective channel gain magnitude.

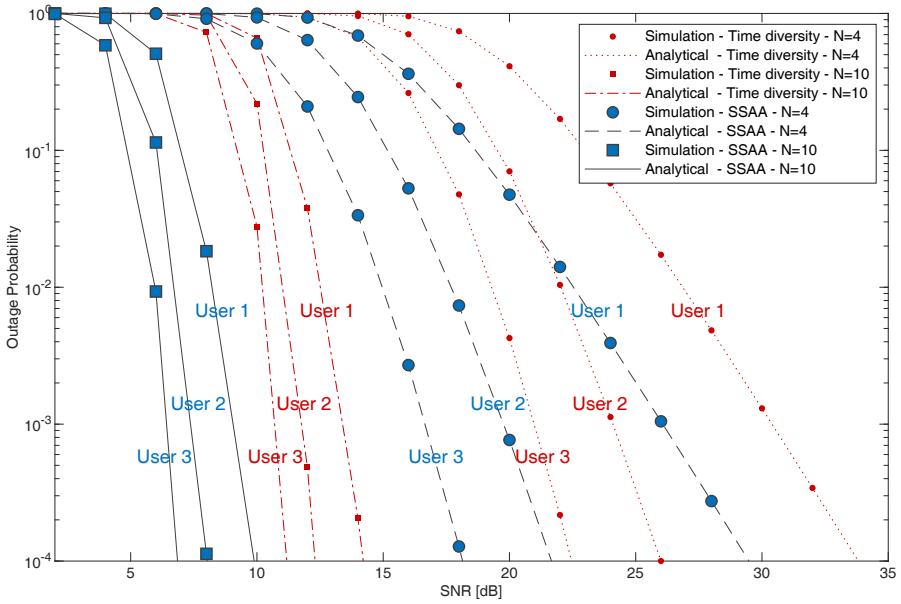
MIMO-OMA with time diversity: For this conventional OMA time diversity scheme, we suppose that there is only one user per group, what means that $G = U$. Aiming to provide fairness in the performance comparisons, the precoding and detection matrices are constructed identically as in the MIMO-NOMA time diversity system, where each user selects its desired symbol from the best of the T receptions. In addition, the total transmit power available for each group is entirely allocated to the only existing user, that is, the power allocation coefficient for each user in the MIMO-OMA system is adjusted to unity.

MIMO-NOMA and MIMO-OMA without diversity: These two schemes consist of conventional full array implementations, the MIMO-NOMA system from [7] and the MIMO-OMA counterpart. In contrast to the time diversity schemes, for these implementations, each distinct symbol is transmitted only one time, so no diversity is explored.

4.5 Numerical and Simulation Results

In this section, some illustrative numerical examples of the proposed system operating with the SSAA technique are presented and compared with the conventional massive MIMO-NOMA and OMA systems described in Section 4.4. The total number of transmit antennas at the BS is set to $M = 90$, in which, for comparison purposes, we adjust the number

Figure 19: Outage probability versus transmit SNR for massive MIMO-NOMA system with SSAA technique and conventional full array system operating in time diversity mode. $Q = T = 3$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



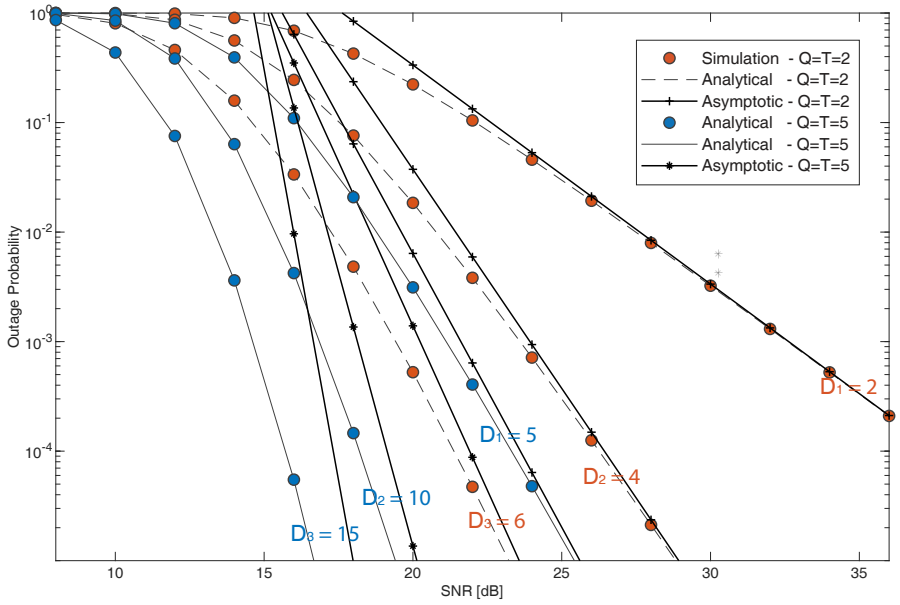
Source: Created by the author.

of sub-arrays in MIMO-NOMA with SSAA to be equal to the number of time slots in the full array time diversity implementation, i.e., $Q = T$. In addition, we consider a scenario with $K = 4$ scattering clusters, where, in each cluster, we assume the existence of 12 users that are further subdivided into $G = 4$ groups with $U = 3$ users each, and we configure the precoders to deliver $\bar{M} = 4$ effective data streams at each transmission for each of the clusters. The angular spread is set to $\Delta_k = 10^\circ$, and the azimuth angle for the k-th cluster is chosen as $\theta_k = \frac{\pi}{4} + \frac{\pi}{2}\left(\frac{k-1}{K-1}\right)$, for $k = 1, \cdots, K$. Furthermore, in order to maximize the array gain, the azimuth angle of the BS is directed to the cluster of interest.

Figure 19 shows the outage probability in terms of transmit SNR. As can be seen, a perfect agreement among simulated and analytical curves is observed. In addition, it can be noticed that the proposed scheme provides remarkable outage performance improvements to massive MIMO-NOMA systems. For example, when employing either $N = 4$ or $N = 10$, all users adopting the SSAA strategy requires roughly 5dB less SNR to achieve the same outage level of that achieved with the full array time diversity scheme, and when $N = 10$, the worst user in SSAA can outperform even the best user in the time diversity counterpart. Figures 20 and 21 bring the validation of the high-SNR analysis derived in Section 4.3.3. One can observe that, for a fixed number of transmit and receive antennas, the system diversity order increases as the number of sub-arrays gets higher, which is in total concordance with the diversity order expression in (58). Moreover, we see that changing the values of target rates and power allocation coefficients does not affect the diversity order of the users, which validates again our analysis.

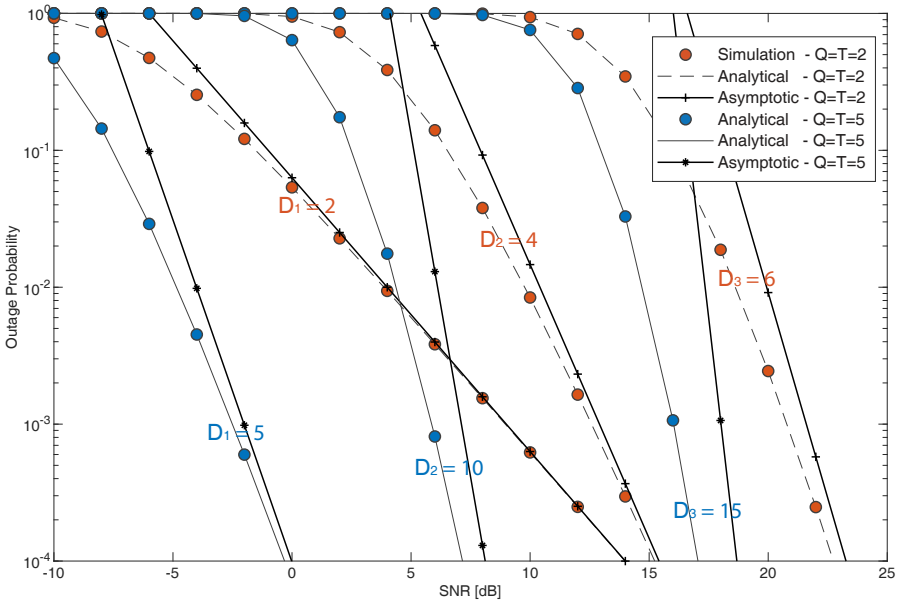
Figure 22 plots the outage sum-rate, $\sum_{u=1}^{U}(1 - P_{k,g,u}^{\text{out}})\mathscr{R}_{u,g,u}$, versus transmit SNR.

Figure 20: Exact and asymptotic outage probability curves for massive MIMO-NOMA system operating with the proposed SSAA technique. $N = 4$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

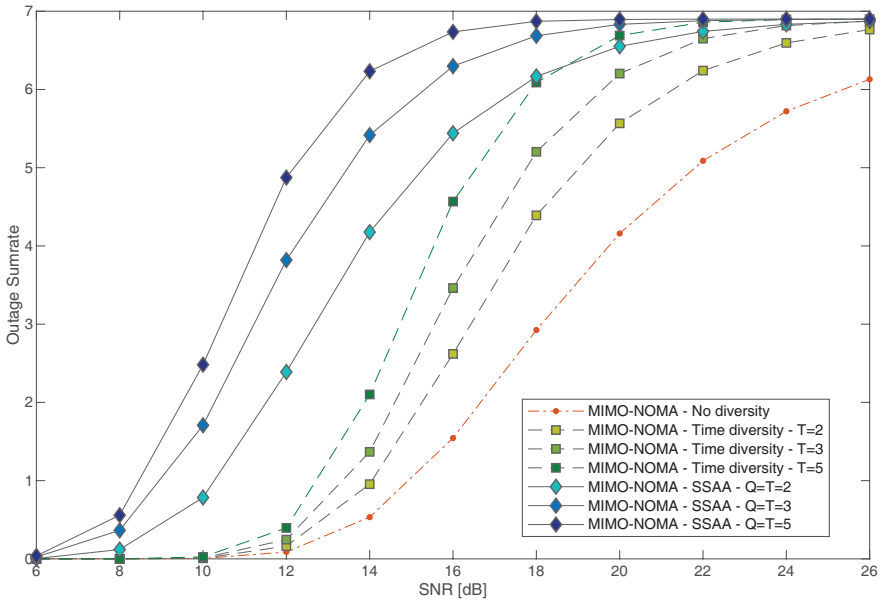
Figure 21: Exact and asymptotic outage probability curves for massive MIMO-NOMA system operating with the proposed SSAA technique. $N = 4$; $\alpha_1^2 = 0.72, \alpha_2^2 = 0.18$, and $\alpha_3^2 = 0.1$; $\mathscr{R}_1 = 0.5, \mathscr{R}_2 = 1$ and $\mathscr{R}_3 = 4.5$ BPCU.



Source: Created by the author.

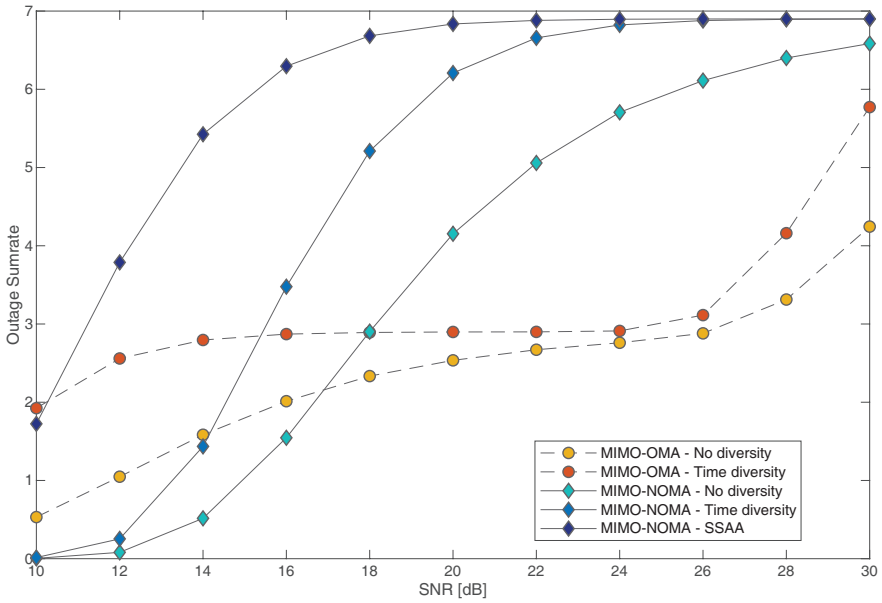
Once again, the benefits that SSAA scheme can provide to massive MIMO-NOMA networks is noticeable. As one can observe, for all values of T, the proposed strategy outperforms the conventional time diversity full array setup. For example, when $T = 3$ and SNR = 12dB, the SSAA scheme achieves an outage sum-rate of 3.76 BPCU against only 0.25 BPCU from the

Figure 22: Outage sum-rate for the proposed SSAA technique and the conventional full array time diversity approach in massive MIMO-NOMA systems. $N = 4$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

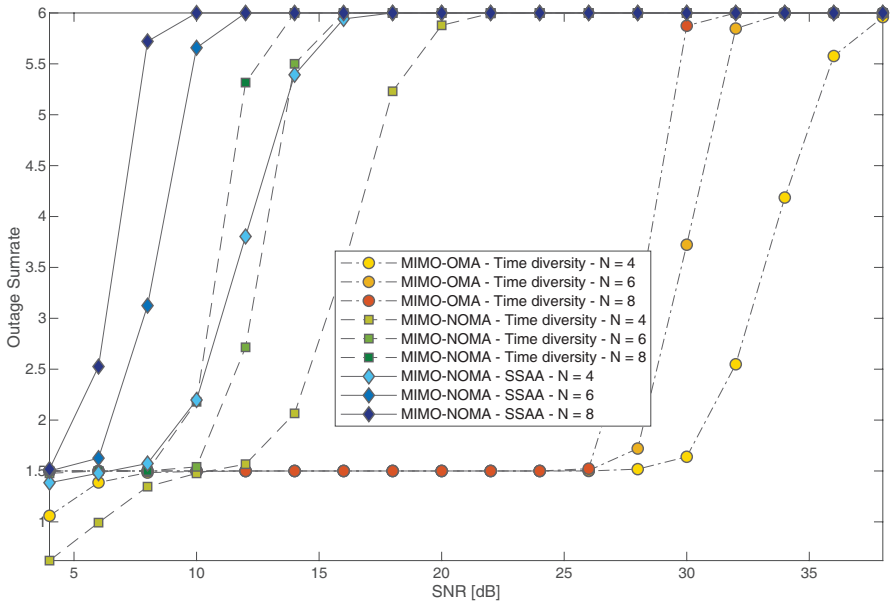
Figure 23: Outage sum-rate for the proposed SSAA technique and various conventional schemes in massive MIMO-NOMA systems. $N = 4$; $Q = T = 3$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

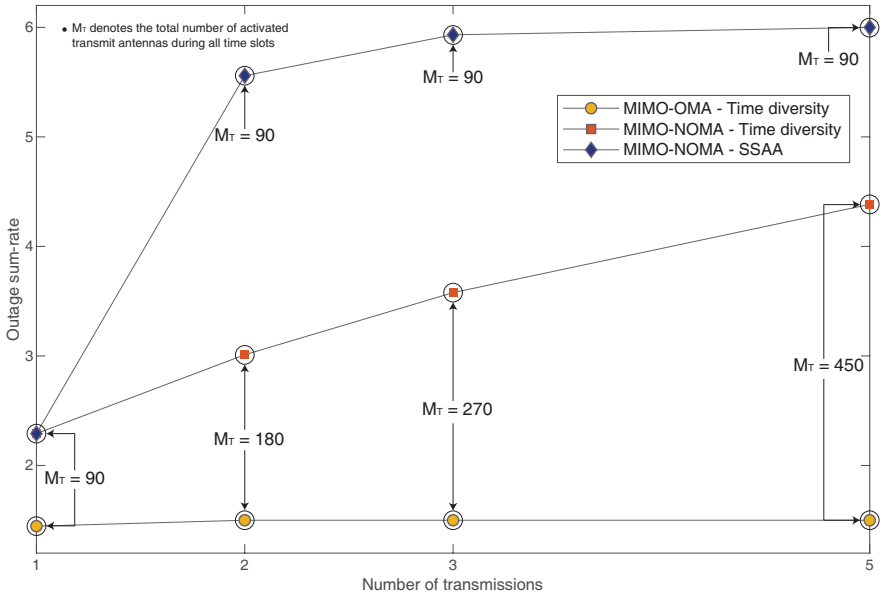
full array system, which represents a spectral gain of almost 15 times. The expressive gain that SSAA can achieve over the system without diversity becomes also evident in this figure, in which a gap of almost 8dB can be observed. In Figure 23, for $Q = T = 3$, the outage sum-rate of the SSAA scheme is compared with conventional full array systems, with and without time diversity. It can be observed that our MIMO-NOMA design with SSAA outperforms all the

Figure 24: Outage sum-rate for the proposed SSAA technique and the conventional full array time diversity approach in massive MIMO-NOMA systems. $Q = 3$; $\alpha_1^2 = 0.72, \alpha_2^2 = 0.18$, and $\alpha_3^2 = 0.1$; $\mathscr{R}_1 = 0.5, \mathscr{R}_2 = 1$ and $\mathscr{R}_3 = 4.5$ BPCU.



Source: Created by the author.

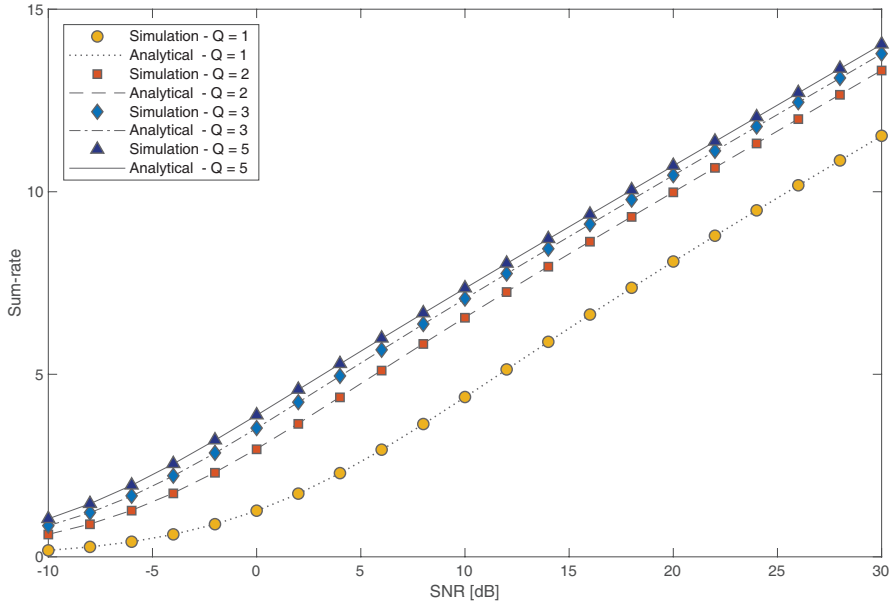
Figure 25: Outage sum-rate versus number of redundant transmissions for a fixed transmit SNR of 16dB. $N = 4$; $\alpha_1^2 = 0.72, \alpha_2^2 = 0.18$, and $\alpha_3^2 = 0.1$; $\mathscr{R}_1 = 0.5, \mathscr{R}_2 = 1$ and $\mathscr{R}_3 = 4.5$ BPCU.



Source: Created by the author.

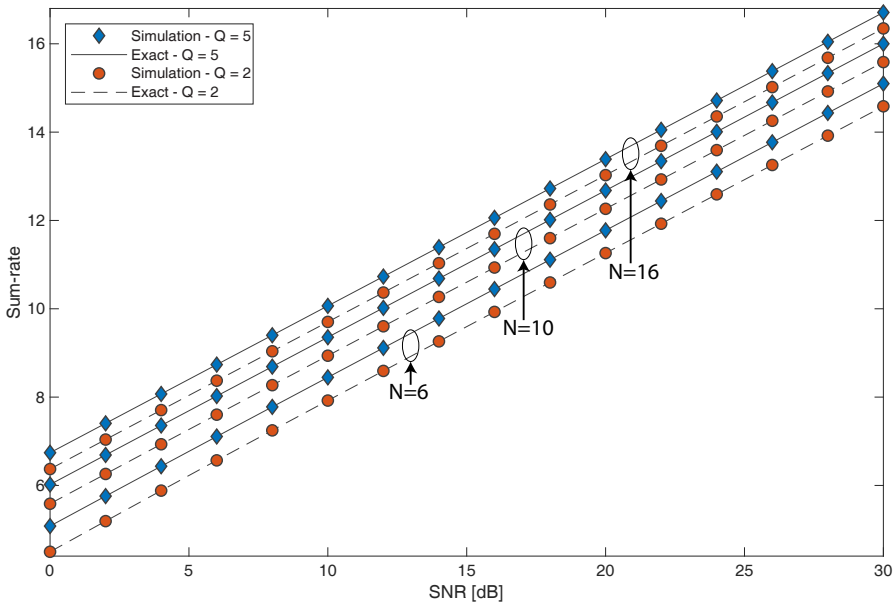
other systems. In particular, considering a transmit SNR of 18dB, the SSAA scheme achieves an outage sum-rate of 6.68 BPCU against 5.21 BPCU of the time diversity MIMO-NOMA system and 2.91 BPCU of the MIMO-NOMA system without diversity. The performance gains over the OMA implementations are even more expressive, reaching up a performance gap of 4.35 BPCU over the MIMO-OMA system without diversity.

Figure 26: Simulated and analytical (generated with (61)) ergodic sum-rate curves versus transmit SNR for massive MIMO-NOMA system operating with the proposed SSAA technique. $N = \bar{M} = 4$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

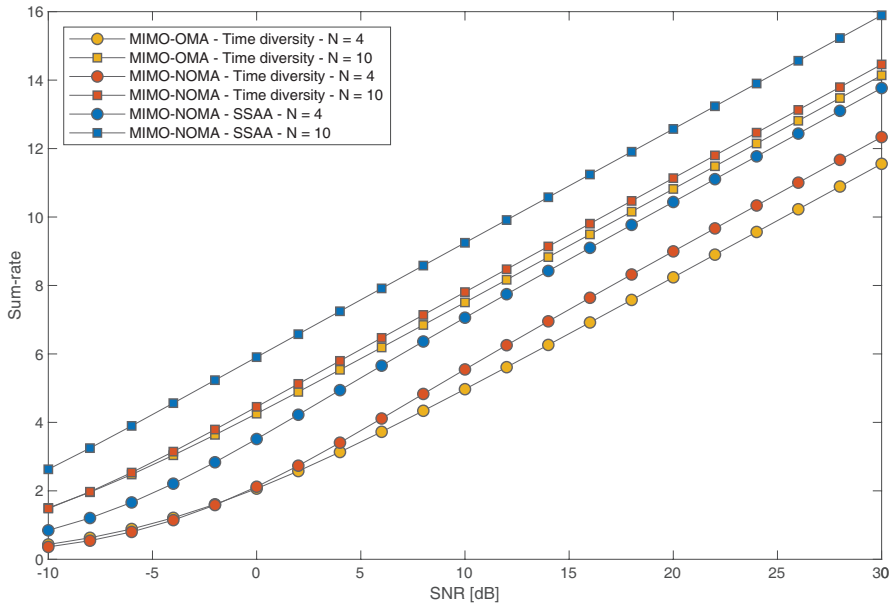
Figure 27: Simulated and exact (generated with (60)) ergodic sum-rate versus transmit SNR for massive MIMO-NOMA system operating with the proposed SSAA technique. $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

Figure 24, brings the outage sum-rate curves for various values of receive antennas considering a different set of power allocation and target rates. As one can notice, the gains achieved by the SSAA scheme over the full array time diversity MIMO-OMA setups are enormous, in which, for all values of N, the curves of the two schemes are separated by more than

Figure 28: Ergodic sum-rate versus transmit SNR for massive MIMO-NOMA with SSAA and the conventional time diversity counterparts. $Q = 3$; $\alpha_1^2 = 0.625, \alpha_2^2 = 0.25$ and $\alpha_3^2 = 0.125$; $\mathscr{R}_1 = 1.4, \mathscr{R}_2 = 1.5$ and $\mathscr{R}_3 = 4$ BPCU.



Source: Created by the author.

20dB. Figure 25 shows the outage sum-rate versus the number of redundant transmissions for a fixed SNR of 16dB. It can be seen again the superior outage sum-rate performance of SSAA for a fixed SNR setup. For instance, with 3 redundant transmissions, the massive MIMO-NOMA system with SSAA achieves a rate of 5.93 BPCU, which is 2.3 BPCU higher than that of the full array MIMO-NOMA system and almost 4 times higher than what is achieved by the MIMO-OMA counterpart. For 5 redundant transmissions, the performance gains of SSAA becomes even more prominent. In addition, as it can be observed, another advantage of our proposed strategy is that, independently of how many retransmissions are performed, the total number of activated antenna elements remains constant, i.e., $M = 90$, what does not happen for the full array schemes.

Figures 26 and 27 validates the analytical development for the ergodic sum-rate in Section 4.3.4. In particular, Figure 26 shows results for various values of Q when $N = \bar{M}$. For this case, the analytical curves are generated with (61), in which a perfect agreement with the simulation results is verified. The cases when $N > \bar{M}$ are shown in Figure 27. In this latter figure, the expression in (60) is used to plot the exact curves. Again, for all values of N and Q, we can observe perfect correspondence among exact and simulated results. Moreover, through both figures, it can be noticed that the system ergodic sum-rate is improved with the increase in the number of sub-arrays. For instance, in Figure 26, for a transmit SNR of 10dB, the system operating with $Q = 5$ sub-arrays achieves an ergodic sum-rate of 7.35 BPCU, which is 2.98 BPCU superior than the achieved with $Q = 1$.

Finally, in Figure 28 the ergodic sum-rate performance of MIMO-NOMA and MIMO-OMA diversity systems are compared, for $N = 4$ and $N = 10$ receive antennas. As it can be

seen, the superiority of the proposed SSAA technique is demonstrated again in terms of ergodic sum-rate. For example, when $N = 10$ and the SNR is 18dB, the MIMO-NOMA system operating with SSAA achieves 1.44 BPCU more sum-rate than the full array time diversity MIMO-NOMA setup and, for $N = 4$, the gap reaches up 1.45 BPCU. If we take into consideration the MIMO-OMA full array system, the gains are even more remarkable.

4.6 Chapter Summary

In this chapter, by partitioning the transmit antennas into multiple sub-arrays at the BS, we have proposed and investigated a novel low-complexity diversity scheme for massive MIMO-NOMA deployments. Detailed design of precoders and detection matrices were presented, and a full in-depth analytical analysis was carried out. In particular, closed-form expressions for the outage probability and ergodic sum-rate were derived. A high SNR asymptotic outage analysis was also conducted, in which the diversity order achieved with the proposed protocol was determined. Furthermore, representative numerical and simulation examples were presented to corroborate the analytical analysis and demonstrate the performance gains of our proposal.

5 CONCLUSIONS AND FUTURE WORKS

Massive MIMO-NOMA has arisen as an essential enabling technology for the ambitious requisites of 5G networks. However, it has been shown that some issues related to physical space constraints and fading channels can limit the performance of such systems and reduce communication reliability. In this regard, aiming to mitigate the impact of such impairments, this dissertation has studied and developed strategies for enhancing the performance of multi-cluster multi-user massive MIMO-NOMA networks under highly correlated scenarios. Specifically, considering the downlink mode, two main original contributions have been accomplished: first, a complete investigation of dual-polarized antennas in massive MIMO-NOMA systems was carried out, and, second, a novel SSAA diversity scheme was proposed for enhancing the performance of each NOMA user. Throughout this work, we have performed full and in-depth analytical studies for both dual-polarized and SSAA MIMO-NOMA designs, where, for each of the systems, we have determined the SINRs, derived exact closed-form expressions for the outage probabilities, and obtained the respective asymptotic approximations in high SNR regime. Then, based on the asymptotic expressions, the diversity gains were evaluated, and, lastly, the ergodic sum-rates were also derived. Furthermore, aiming to validate the analytical analysis and demonstrate the performance superiority of the proposed designs, we have provided representative numerical and simulation results followed by insightful discussions. For instance, in Chapter 3, our results showed that the proposed dual-polarized MIMO-NOMA designs outperform conventional single-polarized systems, even for high cross-polar interference, corroborating the fact that polarization has great potential to bring significant improvements to conventional massive MIMO-NOMA deployments. In the results of Chapter 4, the proposed SSAA scheme outperformed conventional full array massive MIMO-OMA and MIMO-NOMA systems operating with and without the exploration of time diversity. Besides, it became clear the superiority of SSAA in terms of energy consumption, implementation complexity, latency, and feedback overhead, which makes our proposal attractive for applications that require an enhanced performance but have limited resources and restricted computational capabilities.

5.1 Directions for Future Works

The MIMO-NOMA technology is in its infancy, and many open issues still need to be addressed. Consequently, the current work has several interesting potential extensions, in which some of them are highlighted in the following. For example, in this dissertation, we have considered that the BS has perfect knowledge about the user ordering information and that SIC is performed without introducing errors, which are idealistic assumptions. Therefore, the analysis of our proposals under imperfect SIC and imperfect channel state information is one possible future direction. Also, since we have considered only fixed power allocation, another promising possibility would be the development and investigation of more sophisticated

dynamic allocation strategies. For the dual-polarized designs with approach 2, the consideration of high cross-polar interference scenarios would also be interesting to analyze. Regarding the SSAA scheme, even though we have chosen to transmit the symbol replicas in different instants of time, it is possible to separate the transmissions by exploring different domains, such as frequency or code, which arises as potential future work. The combination of concepts of space-time coding and SSAA is another attractive research direction.

REFERENCES

1 REPORT ITU-R M.2370-0: IMT traffic estimates for the years 2020 to 2030. **International Telecommunication Union (ITU)**, Jul. 2015.

2 AKPAKWU, G. A. et al. A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges. **IEEE Access**, v. 6, p. 3619–3647, 2018.

3 DE ALMEIDA, I. B. F. et al. 5G Waveforms for IoT Applications. **IEEE Communications Surveys and Tutorials**, p. 1–1, 2019.

4 LIU, G. et al. 5G features from operation perspective and fundamental performance validation by field trial. **China Communications**, v. 15, n. 11, p. 33–50, Nov. 2018.

5 AL., Y. C. E. Toward the standardization of non-orthogonal multiple access for next generation wireless networks. **IEEE Communications Magazine**, v. 56, n. 3, p. 19–27, Mar. 2018.

6 DING, Z.; ADACHI, F.; POOR, H. V. The application of MIMO to non-orthogonal multiple access. **IEEE Transactions on Wireless Communications**, v. 15, n. 1, p. 537–552, Jan. 2016.

7 DING, Z.; POOR, V. Design of massive-MIMO-NOMA with limited feedback. **IEEE Signal Processing Letters**, v. 23, n. 5, May 2016.

8 AL-HUSSAIBI, W. A.; ALI, F. H. Efficient user clustering, receive antenna selection, and power allocation algorithms for massive MIMO-NOMA systems. **IEEE Access**, v. 7, p. 31865–31882, Feb. 2019.

9 LIEN, S. et al. 5G New Radio: Waveform, Frame Structure, Multiple Access, and Initial Access. **IEEE Communications Magazine**, v. 55, n. 6, p. 64–71, Jun. 2017.

10 WU, X.; BEAULIEU, N. C.; LIU, D. On Favorable Propagation in Massive MIMO Systems and Different Antenna Configurations. **IEEE Access**, v. 5, p. 5578–5593, Apr. 2017.

11 NGO, H. Q.; LARSSON, E. G.; MARZETTA, T. L. Aspects of favorable propagation in Massive MIMO. In: **2014 22nd European Signal Processing Conference (EUSIPCO)**. [S.l.: s.n.], 2014. p. 76–80.

12 JIN, S. et al. On Massive MIMO Zero-Forcing Transceiver Using Time-Shifted Pilots. **IEEE Transactions on Vehicular Technology**, v. 65, n. 1, p. 59–74, Jan. 2016.

13 LARSSON, E. G. et al. Massive MIMO for next generation wireless systems. **IEEE Communications Magazine**, v. 52, n. 2, p. 186–195, Feb. 2014.

14 WANG, X. et al. Millimeter wave communication: A comprehensive survey. **IEEE**

Communications Surveys and Tutorials, v. 20, n. 3, p. 1616–1653, Jun. 2018.

15 NAM, W. et al. Advanced interference management for 5G cellular networks. **IEEE Communications Magazine**, v. 52, n. 5, p. 52–60, May 2014.

16 CHEN, X. et al. Massive MIMO Beamforming With Transmit Diversity for High Mobility Wireless Communications. **IEEE Access**, v. 5, p. 23032–23045, Oct. 2017.

17 LI, Y.; ARUMA BADUGE, G. A. NOMA-Aided Cell-Free Massive MIMO Systems. **IEEE Wireless Communications Letters**, v. 7, n. 6, p. 950–953, Dec. 2018.

18 PARK, J.; CLERCKX, B. Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT. **IEEE Transactions on Wireless Communications**, v. 14, n. 05, May 2015.

19 KIM, T. et al. Limited feedback beamforming systems for dual-polarized MIMO channels. **IEEE Transactions on Wireless Communications**, v. 9, n. 11, Nov. 2010.

20 OESTGES, C.; ERCEG, V.; PAULRAJ, A. J. Propagation modeling of MIMO multipolarized fixed wireless channels. **IEEE Transactions on Vehicular Technology**, v. 53, n. 3, p. 644–654, May 2004.

21 CHENG, X.; HE, Y.; GUIZANI, M. 3-D geometrical model for multi-polarized MIMO systems. **IEEE Access**, v. 5, p. 11974–11984, 2017.

22 CHENG, X.; HE, Y. Geometrical model for point-to-point multi-polarized massive MIMO systems. In: **14th International Conference on Wireless and Mobile Computing (IWCMC)**. [S.l.: s.n.], 2018. p. 435–439. ISSN 2376-6506.

23 TSE, D.; VISWANATH, P. **Fundamentals of Wireless Communication**. [S.l.]: Cambridge University Press, 2005. ISBN 9780521845274.

24 NGUYEN, V. et al. Precoder design for signal superposition in MIMO-NOMA multicell networks. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 12, p. 2681–2695, Dec. 2017.

25 ZHANG, D. et al. Performance analysis of non-regenerative massive-MIMO-NOMA relay systems for 5G. **IEEE Transactions on Communications**, v. 65, n. 11, p. 4777–4790, Nov. 2017.

26 LI, Y. et al. Cooperative non-orthogonal multiple access in multiple-input-multiple-output channels. **IEEE Transactions on Wireless Communications**, v. 17, n. 3, p. 2068–2079, Mar. 2018.

27 CHEN, X.; JIA, R.; NG, D. W. K. The application of relay to massive non-orthogonal

multiple access. **IEEE Transactions on Communications**, v. 66, n. 11, p. 5168–5180, Nov. 2018.

28 LIU, Y. et al. Non-orthogonal multiple access in large-scale heterogeneous networks. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 12, p. 2667–2680, Dec. 2017.

29 ZHENG, F. et al. An efficient CSI feedback scheme for dual-polarized massive MIMO. **IEEE Access**, v. 6, p. 23420–23430, 2018.

30 WU, L. et al. Codebook design for LTE-A downlink system. In: **IEEE Vehicular Technology Conference (VTC Fall)**. [S.l.: s.n.], 2011. p. 1–5.

31 KIM, T. et al. Limited feedback beamforming codebook design for dual-polarized MIMO channels. In: **IEEE Global Communications Conference (GLOBECOM)**. [S.l.: s.n.], 2008. p. 1–5.

32 SU, X. et al. Inter-beam interference cancellation and physical layer security constraints by 3D polarized beamforming in power domain NOMA systems. **IEEE Transactions on Sustainable Computing**, p. 1–1, 2018. ISSN 2377-3782.

33 GONG, M.; YANG, Z.; LYU, B. Antenna Diversity for Downlink MIMO-NOMA Systems With Partial Channel State Information. **IEEE Communications Letters**, v. 22, n. 10, p. 2172–2175, Oct. 2018.

34 GONG, M.; YANG, Z. The application of antenna diversity to NOMA with statistical channel state information. **IEEE Transactions on Vehicular Technology**, v. 68, n. 4, p. 3755–3765, Apr. 2019.

35 TOKA, M.; KUCUR, O. Non-orthogonal multiple access with Alamouti space-time block coding. **IEEE Communications Letters**, v. 22, n. 9, p. 1954–1957, Sep. 2018.

36 PAN, Z. et al. Multi-dimensional Space Time Block Coding Aided Downlink MIMO-SCMA. **IEEE Transactions on Vehicular Technology**, p. 1–1, 2019.

37 KADER, M. F.; SHIN, S. Y. Cooperative relaying using space-time block coded non-orthogonal multiple access. **IEEE Transactions on Vehicular Technology**, v. 66, n. 7, p. 5894–5903, Jul. 2017.

38 ZHAO, J. et al. Dual Relay Selection for Cooperative NOMA With Distributed Space Time Coding. **IEEE Access**, v. 6, p. 20440–20450, 2018.

39 AHMED, H. R.; SOUROUR, E.; ELKAMCHOUCHI, H. M. Analysis for NOMA-CoMP-JT global precoding matrix and IRC receiver for LTE-A. In: **2016 IEEE 13th International Conference on Networking, Sensing, and Control (ICNSC)**. [S.l.: s.n.], 2016. p. 1–6.

40 WANG, H.; LEUNG, S.; SONG, R. Precoding Design for Two-Cell MIMO-NOMA Uplink With CoMP Reception. **IEEE Communications Letters**, v. 22, n. 12, p. 2607–2610, Dec. 2018.

41 YU, Y. et al. Antenna selection for MIMO nonorthogonal multiple access systems. **IEEE Transactions on Vehicular Technology**, v. 67, n. 4, p. 3158–3171, Apr. 2018. ISSN 0018-9545.

42 KHAN, A. S. et al. Network-Coded NOMA With Antenna Selection for the Support of Two Heterogeneous Groups of Users. **IEEE Transactions on Wireless Communications**, v. 18, n. 2, p. 1332–1345, Feb. 2019.

43 DING, Z. et al. A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 10, p. 2181–2195, Oct. 2017.

44 ISLAM, S. M. R. et al. Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges. **IEEE Communications Surveys Tutorials**, v. 19, n. 2, p. 721–742, Oct. 2017.

45 LIU, Y. et al. Nonorthogonal Multiple Access for 5G and Beyond. **Proceedings of the IEEE**, v. 105, n. 12, p. 2347–2381, Dec 2017.

46 DING, Z. et al. Application of Non-Orthogonal Multiple Access in LTE and 5G Networks. **IEEE Communications Magazine**, v. 55, n. 2, p. 185–191, Feb. 2017.

47 DAI, L. et al. Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends. **IEEE Communications Magazine**, v. 53, n. 9, p. 74–81, Sep. 2015.

48 SHIN, W. et al. Non-Orthogonal Multiple Access in Multi-Cell Networks: Theory, Performance, and Practical Challenges. **IEEE Communications Magazine**, v. 55, n. 10, p. 176–183, Oct. 2017.

49 ZHANG, R.; HANZO, L. A unified treatment of superposition coding aided communications: Theory and practice. **IEEE Communications Surveys Tutorials**, v. 13, n. 3, p. 503–520, Jul. 2011.

50 VANKA, S. et al. Superposition coding strategies: Design and experimental evaluation. **IEEE Transactions on Wireless Communications**, v. 11, n. 7, p. 2628–2639, Jul. 2012.

51 KIM, D. et al. Superposition of broadcast and unicast in wireless cellular systems. **IEEE Communications Magazine**, v. 46, n. 7, p. 110–117, Jul. 2008.

52 WANG, L. et al. A comparison of superposition coding schemes. In: **2013 IEEE**

International Symposium on Information Theory. [S.l.: s.n.], 2013. p. 2970–2974.

53 TABASSUM, H. et al. Uplink Vs. Downlink NOMA in Cellular Networks: Challenges and Research Directions. In: **2017 IEEE 85th Vehicular Technology Conference (VTC Spring)**. [S.l.: s.n.], 2017. p. 1–7.

54 ANDREWS, J. G. Interference cancellation for cellular systems: a contemporary overview. **IEEE Wireless Communications**, v. 12, n. 2, p. 19–29, Apr. 2005.

55 MIRIDAKIS, N. I.; VERGADOS, D. D. A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems. **IEEE Communications Surveys Tutorials**, v. 15, n. 1, p. 312–335, Apr. 2013.

56 TELATAR, E. Capacity of multi-antenna gaussian channel. **Bell Laboratories Technical Report**, 1995.

57 FOSCHINI, G. J. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. **Bell Labs Technical Journal**, v. 1, n. 2, p. 41–59, 1996.

58 FOSCHINI, G. J.; GANS, M. J. On limits of wireless communications in a fading environment when using multiple antennas. **Wireless Personal Communications**, v. 6, n. 3, p. 311–335, Mar. 1998.

59 MARZETTA, T. L. Noncooperative cellular wireless with unlimited numbers of base station antennas. **IEEE Transactions on Wireless Communications**, v. 9, n. 11, p. 3590–3600, Nov. 2010.

60 ERICSSON. **Going Massive with MIMO**. 2018. Available in: <https://www.ericsson.com/en/news/2018/1/massive-mimo-highlights>. Accessed on: Jul. 9, 2019.

61 HOYDIS, J.; TEN BRINK, S.; DEBBAH, M. Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need? **IEEE Journal on Selected Areas in Communications**, v. 31, n. 2, p. 160–171, Feb. 2013.

62 NGO, H. Q.; LARSSON, E. G.; MARZETTA, T. L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. **IEEE Transactions on Communications**, v. 61, n. 4, p. 1436–1449, Apr. 2013.

63 NGO, H. Q.; LARSSON, E. G.; MARZETTA, T. L. Aspects of favorable propagation in Massive MIMO. In: **2014 22nd European Signal Processing Conference (EUSIPCO)**. [S.l.: s.n.], 2014. p. 76–80.

64 ROY, M. et al. MIMO Channel Hardening for Ray-based Models. In: **2018 14th**

International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). [S.l.: s.n.], 2018. p. 1–7.

65 TATARIA, H. et al. Spatial correlation variability in multiuser systems. In: **2018 IEEE International Conference on Communications (ICC)**. [S.l.: s.n.], 2018. p. 1–7.

66 CHIZHIK, D. et al. Effect of antenna separation on the capacity of blast in correlated channels. **IEEE Communications Letters**, v. 4, n. 11, p. 337–339, Nov. 2000.

67 SANGUINETTI, L.; BJÖRNSON, E.; HOYDIS, J. Towards Massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination. **arXiv e-prints**, Apr. 2019.

68 BJORNSON, E.; HOYDIS, J.; SANGUINETTI, L. **Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency**. [S.l.]: NOW, 2017. ISBN 9781680839852.

69 POUTANEN, J. et al. Significance of common scatterers in multi-link indoor radio wave propagation. In: **Proceedings of the Fourth European Conference on Antennas and Propagation**. [S.l.: s.n.], 2010. p. 1–5.

70 GE, X. et al. Multi-User Massive MIMO Communication Systems Based on Irregular Antenna Arrays. **IEEE Transactions on Wireless Communications**, v. 15, n. 8, p. 5287–5301, Aug. 2016.

71 SHIU, D. et al. Fading correlation and its effect on the capacity of multielement antenna systems. **IEEE Transactions on Communications**, v. 48, n. 3, p. 502–513, Mar. 2000.

72 ADHIKARY, A. et al. Joint spatial division and multiplexing - The large-scale array regime. **IEEE Transactions on Information Theory**, v. 59, n. 10, p. 6441–6463, Oct. 2013.

73 CLERCKX, B.; OESTGES, C. **MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems**. 2. ed. Orlando, FL, USA: Academic Press, Inc., 2013. ISBN 9780123850553.

74 RUSEK, F. et al. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. **IEEE Signal Processing Magazine**, v. 30, n. 1, p. 40–60, Jan. 2013.

75 ALI, S.; HOSSAIN, E.; KIM, D. I. Non-Orthogonal Multiple Access (NOMA) for Downlink Multiuser MIMO Systems: User Clustering, Beamforming, and Power Allocation. **IEEE Access**, v. 5, p. 565–577, Dec. 2017.

76 RONGTAO XU; LAU, F. C. M. Degradation on the performance of MIMO system under a correlated sub-channels condition. In: **2005 IEEE International Symposium on Circuits and Systems**. [S.l.: s.n.], 2005. v. 5, p. 4967–4970.

77 YANG, S.; HANZO, L. Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs. **IEEE Communications Surveys Tutorials**, v. 17, n. 4, p. 1941–1988, Sep. 2015.

78 AL., F. B. E. Multiple-antenna techniques in LTE-advanced. **IEEE Communications Magazine**, v. 50, n. 3, p. 114–121, Mar. 2012.

79 LIM, C. et al. Recent trend of multiuser MIMO in LTE-advanced. **IEEE Communications Magazine**, v. 51, n. 3, p. 127–135, Mar. 2013.

80 GRAMI, A. **Introduction to Digital Communications**. Boston: Academic Press, 2016. 493 - 527 p. ISBN 9780124076822.

81 GAO, Y.; VINCK, H.; KAISER, T. Massive MIMO Antenna Selection: Switching Architectures, Capacity Bounds, and Optimal Antenna Selection Algorithms. **IEEE Transactions on Signal Processing**, v. 66, n. 5, p. 1346–1360, Mar. 2018.

82 DAVID, H. A.; NAGARAJA, H. N. **Order Statistics**. third. [S.l.]: Wiley Series in Probability and Statistics, 2003.

83 KRISHNAIAH, P. R.; RAO, M. M. Remarks on a multivariate Gamma distribution. **The American Mathematical Monthly**, v. 68, n. 9, p. 342–346, Apr. 1961.

84 ABRAMOWITZ, M.; STEGUN, I. **Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables**. tenth. [S.l.]: USA: National Bureau of Standards, 1972. (55).

85 BATEMAN, H.; ERDELYI, A. **Higher transcendental functions**. [S.l.]: California Institute of Technology Bateman Manuscript Project, New York: McGraw-Hill, 1953-1955, 1955. v. 2.

86 GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of Integrals, Series, and Products**. Seventh. [S.l.]: Academic Press, 2007.

APPENDIX A – PROOF OF LEMMA 3.1

To successfully apply SIC, the effective channel gains of the users within a group must be ordered. Then, assuming that $\zeta_{k,g,1} < \zeta_{k,g,2} < \cdots < \zeta_{k,g,U}$, (23) can be rewritten as

$$\hat{s}_{k,g,u} = \underbrace{\alpha_{k,g,u} s_{k,g,u}}_{\text{signal of interest}} + \underbrace{\sum_{n=k+1}^{U} \alpha_{k,g,n} s_{k,g,n}}_{\text{interference}} + \underbrace{[\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u}]_g}_{\text{noise}}. \tag{A-1}$$

From (A-1), the message intended to the i-th weaker user that is observed at the u-th user within the g-th group of the k-th cluster is decoded with the following SNR

$$\gamma^{i}_{k,g,u} = \frac{\mathbb{E}[|\alpha_{k,g,i} s_{k,g,i}|^2]}{\mathbb{E}\left[\sum_{j=i+1}^{U} |\alpha_{k,g,j} s_{k,g,j}|^2\right] + \mathbb{E}[|[\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u}]_g|^2]}. \tag{A-2}$$

Knowing that

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u}\|^2] &= \mathbb{E}[\mathrm{tr}\{(\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u})(\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u})^H\}] = \mathrm{tr}\{\sigma_n^2 \mathbf{I}_N \mathbb{E}[\bar{\mathbf{H}}^{\dagger H}_{k,g,u} \bar{\mathbf{H}}^{\dagger}_{k,g,u}]\} \\ &= \sigma_n^2 \mathbb{E}[\mathrm{tr}\{\bar{\mathbf{H}}^{\dagger}_{k,g,u} \bar{\mathbf{H}}^{\dagger H}_{k,g,u}\}], \end{aligned} \tag{A-3}$$

and since only the g-th element of $\hat{\mathbf{s}}_{k,g,u}$ is desired, the second term in the denominator of (A-2) can be simplified to

$$\mathbb{E}[|[\bar{\mathbf{H}}^{\dagger}_{k,g,u} \mathbf{n}_{k,g,u}]_g|^2] = \sigma_n^2 \mathbb{E}[\mathrm{tr}\{[\bar{\mathbf{H}}^{\dagger}_{k,g,u} \bar{\mathbf{H}}^{\dagger H}_{k,g,u}]_{g,g}\}] = \sigma_n^2 \|[\bar{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2, \tag{A-4}$$

where σ_n^2 is the variance of the noise vector. Now, by substituting (A-4) in (A-2) and defining the transmit SNR as $\rho = \frac{1}{\sigma_n^2}$, we get

$$\gamma^{i}_{k,g,u} = \frac{\frac{1}{\|[\bar{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2} \alpha^2_{k,g,i}}{\sum_{j=i+1}^{U} \frac{1}{\|[\bar{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2} \alpha^2_{k,g,j} + \frac{1}{\rho}}, \qquad \text{for} \quad 1 \le i \le u < U. \tag{A-5}$$

From (A-5), the effective channel gain can be defined as $\zeta_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2}$. In addition, it can be noticed that when $i = u = U$, there is no interfering messages and the first term of the denominator in (A-5) equals zero. Then, the following is defined

$$\mathscr{I}_i = \begin{cases} \sum_{j=i+1}^{U} \alpha^2_{k,g,j}, & \text{for} \quad 1 \le i \le u < U, \\ 0, & \text{for} \quad i = u = U. \end{cases} \tag{A-6}$$

Finally, by replacing $\zeta_{k,g,u}$ and (A-6) in (A-5), the SINR expression becomes

$$\gamma^{j}_{k,g,u} = \frac{\zeta_{k,g,u}\alpha^{2}_{k,g,i}}{\zeta_{k,g,u}\mathscr{I}_{i} + \frac{1}{\rho}}, \qquad \text{for} \quad 1 \leq i \leq U, \tag{A-7}$$

which concludes the proof of Lemma 3.1.

APPENDIX B – PROOF OF LEMMA 3.2

Similarly to Appendix A, we consider that the effective channel gains are sorted out in increasing order, i.e., $\zeta^*_{k,g,1} < \zeta^*_{k,g,2} < \cdots < \zeta^*_{k,g,U}$. Then, (28) can be rearranged as

$$\hat{s}_{k,g,u} = \underbrace{\alpha_{k,g,u} s_{k,g,u}}_{\text{signal of interest}} + \underbrace{\sum_{n=k+1}^{U} \alpha_{k,g,n} s_{k,g,n}}_{\text{interference}} + \underbrace{[\bar{\mathbf{H}}^{\dagger p}_{k,g,u} \mathbf{n}^{p}_{k,g,u}]_g}_{\text{noise}}, \tag{B-1}$$

where $p \in \{v,h\}$ denotes the polarization with maximum effective channel gain. From (B-1), the SINR that is obtained at the u-th user while decoding the i-th message is given by

$$\gamma^{i}_{k,g,u} = \frac{\mathbb{E}[|\alpha_{k,g,i} s_{k,g,i}|^2]}{\mathbb{E}\left[\sum_{j=i+1}^{U} |\alpha_{k,g,j} s_{k,g,j}|^2\right] + \mathbb{E}[|[\bar{\mathbf{H}}^{\dagger p}_{k,g,u} \mathbf{n}^{p}_{k,g,u}]_g|^2]} = \frac{\frac{1}{\|[\bar{\mathbf{H}}^{\dagger p}_{k,g,u}]_{g,*}\|^2} \alpha^2_{k,g,i}}{\sum_{j=i+1}^{U} \frac{1}{\|[\bar{\mathbf{H}}^{\dagger p}_{k,g,u}]_{g,*}\|^2} \alpha^2_{k,g,j} + \frac{1}{\rho}}. \tag{B-2}$$

Since p represents the polarization that provides the best channel quality, one can define $\zeta^{v}_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger v}_{k,g,u}]_{g,*}\|^2}$ and $\zeta^{h}_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger h}_{k,g,u}]_{g,*}\|^2}$. As a result, the effective channel gain for the u-th user can be denoted by $\zeta^*_{k,g,u} = \max\{\zeta^{v}_{k,g,u}, \zeta^{h}_{k,g,u}\}$. Thus, (B-2) can be rewritten as

$$\gamma^{i}_{k,g,u} = \frac{\zeta^*_{k,g,u} \alpha^2_{k,g,i}}{\zeta^*_{k,g,u} \mathscr{I}_i + \frac{1}{\rho}}, \qquad \text{for} \quad 1 \leq i \leq u \leq U. \tag{B-3}$$

where ρ and $\mathscr{I}_i$ have the same definition as in Appendix A. This completes the proof.

APPENDIX C – PROOF OF PROPOSITION 3.1

We first simplify (55) as following

$$P^{\text{out}}_{k,g,u} = P[\log_2(1+\gamma^{j}_{k,g,u}) < \mathscr{R}_{k,g,i}] = P[\gamma^{j}_{k,g,u} < 2^{\mathscr{R}_{k,g,i}} - 1]. \tag{C-1}$$

Then, by replacing (53) in (C-1) and applying simple algebraic manipulations, we get

$$P^{\text{out}}_{k,g,u} = P\left[\zeta_{k,g,u} < \frac{1}{\rho}\frac{2^{\mathscr{R}_{k,g,i}} - 1}{\alpha^2_{k,g,i} - \mathscr{I}_i(2^{\mathscr{R}_{k,g,i}} - 1)}\right]. \tag{C-2}$$

Let $\Upsilon_{k,g,u} = \max\limits_{i\in[1,u]}\left\{\frac{2^{\mathscr{R}_{k,g,i}}-1}{\alpha^2_{k,g,i}-\mathscr{I}_i(2^{\mathscr{R}_{k,g,i}}-1)}\right\}$. Then, (C-2) becomes

$$P^{\text{out}}_{k,g,u} = P\left[\zeta_{k,g,u} < \rho^{-1}\Upsilon_{k,g,u}\right]. \tag{C-3}$$

Note that (C-3) represents the cumulative distribution function (CDF) of the effective channel gain. Thus, in what follows, the statistical properties of $\zeta_{k,g,u}$ must be identified. As established in Lemma I, the effective channel gain for the u-th user is given by $\zeta_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}^{\dagger}_{k,g,u}]_{g,*}\|^2}$. This value corresponds to the inverse g-th diagonal element of the covariance matrix $\boldsymbol{\Xi}_{k,g,u} = \bar{\mathbf{H}}^{\dagger}_{k,g,u}\bar{\mathbf{H}}^{\dagger H}_{k,g,u}$. Thus, with the purpose of determining the main diagonal marginal distribution of $\boldsymbol{\Xi}_{k,g,u}$, we perform the following expansion

$$\begin{aligned}\boldsymbol{\Xi}_{k,g,u} &= (\bar{\mathbf{H}}^{H}_{k,g,u}\bar{\mathbf{H}}_{k,g,u})^{-1}\bar{\mathbf{H}}^{H}_{k,g,u}\bar{\mathbf{H}}_{k,g,u}(\bar{\mathbf{H}}^{H}_{k,g,u}\bar{\mathbf{H}}_{k,g,u})^{-1}\\ &= \frac{1}{\chi+1}\mathbf{I}_2 \otimes (\tilde{\mathbf{B}}^{H}_{k}\tilde{\mathbf{R}}_k\tilde{\mathbf{B}}_k)^{-1} = \frac{1}{\chi+1}\mathbf{I}_2 \otimes \tilde{\mathbf{W}}^{-1}_{k} = \frac{1}{\chi+1}\mathbf{W}^{-1}_{k},\end{aligned} \tag{C-4}$$

where $\tilde{\mathbf{W}}_k = \tilde{\mathbf{B}}^{H}_{k}\tilde{\mathbf{R}}_k\tilde{\mathbf{B}}_k$ and $\mathbf{W}^{-1}_{k} = \mathbf{I}_2 \otimes \tilde{\mathbf{W}}^{-1}_{k}$. Given that the matrices $\mathbf{G}^{i\to j}_{k,l}$ have complex Gaussian entries, the resulting matrix $\mathbf{G}^{i\to j}_{k,l}(\mathbf{G}^{i\to j}_{k,l})^H$ is complex Wishart distributed. As a consequence, the matrix $\tilde{\mathbf{W}}_k$ also follows the Wishart distribution and, in this way, $\tilde{\mathbf{W}}^{-1}_{k}$ is inverse Wishart distributed with $\frac{N}{2}$ degrees of freedom and covariance matrix $(\tilde{\mathbf{B}}^{H}_{k}\tilde{\mathbf{R}}_k\tilde{\mathbf{B}}_k)^{-1}$, that is, $\tilde{\mathbf{W}}^{-1}_{k} \sim \mathscr{W}^{-1}_{\frac{\tilde{M}}{2}}(\frac{N}{2}, (\tilde{\mathbf{B}}^{H}_{k}\tilde{\mathbf{R}}_k\tilde{\mathbf{B}}_k)^{-1})$[7, 82].

At this point, we already know the distribution of $\tilde{\mathbf{W}}^{-1}_{k}$. However, we are interested in the distribution of the inverse of the main diagonal elements of $\boldsymbol{\Xi}_{k,g,u}$. The marginal distribution of the main diagonal of an inverse Wishart matrix also follows the inverse Wishart distribution and it is equivalent to the univariate inverse Gamma distribution [82, 83]. Thus, the inverse of the diagonal elements of $\mathbf{W}^{-1}_{k}$ follows the Gamma distribution. Consequently, considering the unordered case, the CDF for the inverse of diagonal elements of $\boldsymbol{\Xi}_{k,g,u}$ is given

by [84]

$$F(x) = \frac{\gamma\big(\eta, (\chi+1)\beta_{k,g}x\big)}{\Gamma(\eta)}, \tag{C-5}$$

and the corresponding probability density function (PDF) can be written as

$$f(x) = \frac{(\chi+1)^{\eta}\beta_{k,g}^{\eta}x^{\eta-1}e^{-(\chi+1)\beta_{k,g}x}}{\Gamma(\eta)}, \tag{C-6}$$

where $\eta = \frac{N}{2} - \frac{\bar{M}}{2} + 1$ and $\beta_{k,g} = [\mathbf{W}_k^{-1}]_{g,g} = [\mathbf{I}_2 \otimes (\tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k)^{-1}]_{g,g}$. For the ordered case, we can consider the effective channel gain $\zeta_{k,g,u}$ as the u-th order statistic. Therefore, the PDF for the u-th ordered channel gain can be obtained as [82]

$$f_{\zeta_{k,g,u}}(x) = {}_U\binom{U-1}{u-1} f(x)F(x)^{u-1}(1-F(x))^{U-u}. \tag{C-7}$$

Rewriting the term $(1-F(x))^{U-u}$ as $\sum_{n=0}^{U-u}(-1)^n\binom{U-u}{n}F(x)^n$ and defining $\mathscr{U}_u = U\binom{U-1}{u-1}$, (C-7) can be simplified as

$$\begin{aligned} f_{\zeta_{k,g,u}}(x) &= \mathscr{U}_u \sum_{n=0}^{U-u} {}_{(-1)^n}\binom{U-u}{n} f(x)F(x)^{u-1+n} \\ &= \mathscr{U}_u \sum_{n=0}^{U-u} {}_{(-1)^n}\binom{U-u}{n} \frac{(\chi+1)^{\eta}\beta_{k,g}^{\eta}}{\Gamma(\eta)^{u+n}} x^{\eta-1}e^{-(\chi+1)\beta_{k,g}x}\gamma\big(\eta,(\chi+1)\beta_{k,g}x\big)^{u-1+n}. \end{aligned} \tag{C-8}$$

Finally, by integrating (C-8), a closed-form expression for the outage probability of the u-th user in the g-th group of the k-cluster can de derived as

$$\begin{aligned} P_{k,g,u}^{\text{out}} &= \mathscr{U}_u \sum_{n=0}^{U-u} {}_{(-1)^n}\binom{U-u}{n} \frac{(\chi+1)^{\eta}\beta_{k,g}^{\eta}}{\Gamma(\eta)^{u+n}} \int_0^{\rho^{-1}\Upsilon_{k,g,u}} x^{\eta-1}e^{-(\chi+1)\beta_{k,g}x}\gamma\big(\eta,(\chi+1)\beta_{k,g}x\big)^{u-1+n}dx \\ &= \mathscr{U}_u \sum_{n=0}^{U-u} {}_{(-1)^n}\binom{U-u}{n} \frac{1}{u+n}\left[\frac{\gamma\big(\eta,(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}\big)}{\Gamma(\eta)}\right]^{u+n}, \end{aligned} \tag{C-9}$$

which completes the proof.

APPENDIX D – PROOF OF PROPOSITION 3.2

The outage probability for the Approach II can be derived from

$$P_{k,g,u}^{\text{out}} = P\left[\zeta_{k,g,u}^{*} < \rho^{-1}\Upsilon_{k,g,u}\right], \tag{D-1}$$

which corresponds to the the CDF of the random variable $\zeta_{k,g,u}^{*} = \max\{\zeta_{k,g,u}^{h}, \zeta_{k,g,u}^{v}\}$. Given this, first we need to perform a statistical characterization of the channel gains of each polarization, $\zeta_{k,g,u}^{h}$ and $\zeta_{k,g,u}^{v}$ and then, later, we can find the distribution functions of $\zeta_{k,g,u}^{*}$.

The channel gains $\zeta_{k,g,u}^{h}$ and $\zeta_{k,g,u}^{v}$ correspond to the inverse of the g-th main diagonal element of the covariance matrices $\boldsymbol{\Xi}_{k,g,u}^{h} = \bar{\mathbf{H}}_{k,g,u}^{\dagger h}(\bar{\mathbf{H}}_{k,g,u}^{\dagger h})^{H}$ and $\boldsymbol{\Xi}_{k,g,u}^{v} = \bar{\mathbf{H}}_{k,g,u}^{\dagger v}(\bar{\mathbf{H}}_{k,g,u}^{\dagger v})^{H}$, respectively. In view of this, we expand the matrix $\boldsymbol{\Xi}_{k,g,u}^{h}$ in the following way

$$\begin{aligned}
\boldsymbol{\Xi}_{k,g,u}^{h} &= ((\bar{\mathbf{H}}_{k,g,u}^{h})^{H}\bar{\mathbf{H}}_{k,g,u}^{h})^{-1}(\bar{\mathbf{H}}_{k,g,u}^{h})^{H}\bar{\mathbf{H}}_{k,g,u}^{h}((\bar{\mathbf{H}}_{k,g,u}^{h})^{H}\bar{\mathbf{H}}_{k,g,u}^{h})^{-1} \\
&= (\tilde{\mathbf{B}}_{k}^{H}\mathbf{H}_{k,g,u}^{h\to h}(\mathbf{H}_{k,g,u}^{h\to h})^{H}\tilde{\mathbf{B}}_{k})^{-1} \\
&= \frac{1}{\chi+1}(\tilde{\mathbf{B}}_{k}^{H}\tilde{\mathbf{R}}_{k}\tilde{\mathbf{B}}_{k})^{-1} \\
&= \frac{1}{\chi+1}\tilde{\mathbf{W}}^{-1},
\end{aligned} \tag{D-2}$$

which is also valid for $\boldsymbol{\Xi}_{k,g,u}^{v}$, since both polarizations share the same covariance matrix.

As discussed in Appendix C, $\tilde{\mathbf{W}}^{-1}$ follows the inverse Wishart distribution, i.e. $\tilde{\mathbf{W}}_{k}^{-1} \sim \mathscr{W}_{\frac{\tilde{M}}{2}}^{-1}(\frac{N}{2}, (\tilde{\mathbf{B}}_{k}^{H}\tilde{\mathbf{R}}_{k}\tilde{\mathbf{B}}_{k})^{-1})$, and the inverse of its main diagonal elements follows the Gamma distribution. Consequently, the effective channel gains $\zeta_{k,g,u}^{h}$ and $\zeta_{k,g,u}^{v}$ can be seen as two independent and identically distributed Gamma random variables. Hence, first considering that the gains are not ordered, the CDF of $\max\{\zeta_{k,g,u}^{h}, \zeta_{k,g,u}^{v}\}$ is given by

$$F_{\max}(x) = F_{\zeta_{k,g,u}^{h}}(x)F_{\zeta_{k,g,u}^{v}}(x) = \frac{\gamma\left(\eta, (\chi+1)\tilde{\beta}_{k,g}x\right)^{2}}{\Gamma(\eta)^{2}}, \tag{D-3}$$

and its PDF by

$$f_{\max}(x) = \frac{\partial F_{\max}(x)}{\partial x} = \frac{2(\chi+1)^{\eta}\tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2}}x^{\eta-1}e^{-(\chi+1)\tilde{\beta}_{k,g}x}\gamma\left(\eta, (\chi+1)\tilde{\beta}_{k,g}x\right), \tag{D-4}$$

where η is defined identically as in Appendix C, and $\tilde{\beta}_{k,g} = [\tilde{\mathbf{W}}_{k}^{-1}]_{g,g} = [(\tilde{\mathbf{B}}_{k}^{H}\tilde{\mathbf{R}}_{k}\tilde{\mathbf{B}}_{k})^{-1}]_{g,g}$.

As a result, the PDF for the ordered effective channel gains $\zeta^*_{k,g,u}$ is obtained by

$$\begin{aligned} f_{\zeta^*_{k,g,u}}(x) &= \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} f_{\max}(x) F_{\max}(x)^{u-1+n} \\ &= \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta} \tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} x^{\eta-1} e^{-(\chi+1)\tilde{\beta}_{k,g}x} \gamma\Big(\eta, (\chi+1)\tilde{\beta}_{k,g}x\Big)^{1+2(u-1+n)}. \end{aligned} \tag{D-5}$$

At last, the closed-form general outage probability expression for approach 2 is obtained by integrating (D-5), as following

$$\begin{aligned} P^{\text{out}}_{k,g,u} =& \ \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta} \tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} \\ & \times \int_0^{\rho^{-1}\Upsilon_{k,g,u}} x^{\eta-1} e^{-(\chi+1)\tilde{\beta}_{k,g}x} \gamma\Big(\eta, (\chi+1)\tilde{\beta}_{k,g}x\Big)^{1+2(u-1+n)} dx \\ =& \ \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta} \tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} \\ & \times \left[\frac{(\chi+1)^{-\eta} \tilde{\beta}_{k,g}^{-\eta} \gamma\Big(\eta, (\chi+1)\tilde{\beta}_{k,g}x\Big)^{2+2(u-1+n)}}{2+2(u-1+n)} \right]_0^{\rho^{-1}\Upsilon_{k,g,u}} \\ =& \ \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{1}{u+n} \left[\frac{\gamma\Big(\eta, (\chi+1)\tilde{\beta}_{k,g}\rho^{-1}\Upsilon_{k,g,u}\Big)}{\Gamma(\eta)} \right]^{2(u+n)}, \end{aligned} \tag{D-6}$$

which completes the proof.

APPENDIX E – PROOF OF PROPOSITION 3.3

Since $\eta = \frac{N}{2} - \frac{\tilde{M}}{2} + 1$, and considering that $\frac{N}{2} \geq \frac{\tilde{M}}{2}$, η assumes only positive integer values. Then, a series representation for this special case can be applied to simplify the Gamma functions presented in (56) and (34), as follows [85, 86]

$$\begin{aligned}\frac{\gamma\left(\eta,(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}\right)}{\Gamma(\eta)} &= \frac{(\eta-1)!\left(1-e^{-(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}}\sum_{m=0}^{\eta-1}\frac{[(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}]^m}{m!}\right)}{(\eta-1)!} \\ &= e^{-(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}}\sum_{m=\eta}^{\infty}\frac{[(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}]^m}{m!} \\ &< \frac{[(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}]^{\eta}}{\eta!}. \end{aligned} \tag{E-1}$$

Thus, the high-SNR outage probability approximations for Approaches I and II can be obtained by replacing (E-1) in (56) and (34), which results in

- Asymptotic Outage Probability for Approach I:

$$\begin{aligned}P_{k,g,u}^{\text{out}} &\approx \mathscr{U}_u\sum_{n=0}^{U-u}(-1)^n\binom{U-u}{n}\frac{1}{u+n}\frac{[(\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u}]^{\eta(u+n)}}{(\eta!)^{u+n}} \\ &\approx \frac{1}{\rho^{\eta u}}\frac{\mathscr{U}_u}{u}\frac{[(\chi+1)\beta_{k,g}\Upsilon_{k,g,u}]^{\eta u}}{(\eta!)^u}. \end{aligned} \tag{E-2}$$

The diversity order reached by the u-th user is

$$\mathscr{D}_u = \eta u = \left(\frac{N}{2} - \frac{\tilde{M}}{2} + 1\right)u. \tag{E-3}$$

- Asymptotic Outage Probability for Approach II:

$$\begin{aligned}P_{k,g,u}^{\text{out}} &\approx \mathscr{U}_u\sum_{n=0}^{U-u}(-1)^n\binom{U-u}{n}\frac{1}{u+n}\frac{[(\chi+1)\tilde{\beta}_{k,g}\rho^{-1}\Upsilon_{k,g,u}]^{\eta(2(u+n))}}{(\eta!)^{2(u+n)}} \\ &\approx \frac{1}{\rho^{2\eta u}}\frac{\mathscr{U}_u}{u}\frac{[(\chi+1)\tilde{\beta}_{k,g}\Upsilon_{k,g,u}]^{2\eta u}}{(\eta!)^{2u}}. \end{aligned} \tag{E-4}$$

The diversity order reached by the u-th user is

$$\mathscr{D}_u = 2\eta u = (N - \bar{M} + 2)\, u. \tag{E-5}$$

This completes the proof.

APPENDIX F – PROOF OF PROPOSITION 3.4

By replacing (53) in (39) and performing some simplifications, we have

$$C_{k,g} = \sum_{u=1}^{U} \log_2 \left(1 + \frac{\rho \zeta_{k,g,u} \alpha_{k,g,u}^2}{\rho \zeta_{k,g,u} \mathscr{I}_u + 1} \right) = \sum_{u=1}^{U} \log_2 \left(\frac{1 + \rho \zeta_{k,g,u} (\alpha_{k,g,u}^2 + \mathscr{I}_u)}{1 + \rho \zeta_{k,g,u} \mathscr{I}_u} \right). \tag{F-1}$$

Then, let $\varepsilon_{1,u} = \rho(\alpha_{k,g,u}^2 + \mathscr{I}_u)$ and $\varepsilon_{2,u} = \rho \mathscr{I}_u$, which yields

$$C_{k,g} = \sum_{u=1}^{U} \log_2 \left(\frac{1 + \zeta_{k,g,u} \varepsilon_{1,u}}{1 + \zeta_{k,g,u} \varepsilon_{2,u}} \right). \tag{F-2}$$

In order to obtain the desired ergodic sum-rate expression, we need to calculate the expectation of the random variable $\log_2 \left(\frac{1+\zeta_{k,g,u}\varepsilon_{1,u}}{1+\zeta_{k,g,u}\varepsilon_{2,u}} \right)$. This can be accomplished from the PDF $\zeta_{k,g,u}$, yielding

$$f_{\zeta_{k,g,u}}(x) = \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{(\chi+1)^\eta \beta_{k,g}^\eta}{\Gamma(\eta)^{u+n}} x^{\eta-1} e^{-(\chi+1)\beta_{k,g} x} \gamma\big(\eta, (\chi+1)\beta_{k,g} x\big)^{u-1+n}. \tag{F-3}$$

Thus, the ergodic sum-rate for the g-th group in the k-th cluster can be evaluated by

$$\begin{aligned} \bar{C}_{k,g} = \sum_{u=1}^{U} \int_0^\infty \log_2 \left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}} \right) f_{\zeta_{k,g,u}}(x) dx = \sum_{u=1}^{U} \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{(\chi+1)^\eta \beta_{k,g}^\eta}{\Gamma(\eta)^{u+n}} \\ \times \int_0^\infty \log_2 \left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}} \right) x^{\eta-1} e^{-(\chi+1)\beta_{k,g} x} \gamma\big(\eta, (\chi+1)\beta_{k,g} x\big)^{u-1+n} dx, \end{aligned} \tag{F-4}$$

which completes the proof.

APPENDIX G – PROOF OF PROPOSITION 3.5

Similarly to the analysis employed to Approach I, the sum-rate expression in (39) can be simplified for Approach 2 as

$$C_{k,g} = \sum_{u=1}^{U} \log_2 \left(\frac{1+\zeta^*_{k,g,u}\varepsilon_{1,u}}{1+\zeta^*_{k,g,u}\varepsilon_{2,u}} \right). \tag{G-1}$$

It can be shown that the PDF of $\zeta^*_{k,g,u}$ can be written as

$$f_{\zeta_{k,g,u}}(x) = \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta}\tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} x^{\eta-1} e^{-(\chi+1)\tilde{\beta}_{k,g}x} \gamma\left(\eta, (\chi+1)\tilde{\beta}_{k,g}x\right)^{1+2(u-1+n)}. \tag{G-2}$$

Consequently, the ergodic sum-rate for the g-th group in the k-th cluster can be derived as

$$\begin{aligned}
\bar{C}_{k,g} &= \sum_{u=1}^{U} \int_0^{\infty} \log_2 \left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}} \right) f_{\zeta^*_{k,g,u}}(x)dx \\
&= \sum_{u=1}^{U} \mathscr{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^{\eta}\tilde{\beta}_{k,g}^{\eta}}{\Gamma(\eta)^{2+2(u-1+n)}} \\
&\times \int_0^{\infty} \log_2 \left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}} \right) x^{\eta-1} e^{-(\chi+1)\tilde{\beta}_{k,g}x} \gamma\left(\eta, (\chi+1)\tilde{\beta}_{k,g}x\right)^{1+2(u-1+n)} dx,
\end{aligned} \tag{G-3}$$

which completes the proof.

APPENDIX H – PROOF OF LEMMA 4.1

From (52), one can see that the u-th user decodes the i-th weaker message, for $1 \leq i \leq u < U$, with the following SINR

$$\begin{aligned}\gamma^{i}_{k,g,u} &= \frac{\mathbb{E}[|\alpha_{k,g,i}s_{k,g,i}|^2]}{\mathbb{E}\left[\sum_{j=i+1}^{U}|\alpha_{k,g,j}x_{k,g,j}|^2\right]+\mathbb{E}[|[\mathbf{H}^{q\dagger}_{k,g,u}\mathbf{n}^{q}_{k,g,u}]_g|^2]} \\ &= \frac{\alpha^2_{k,g,i}}{\sum_{j=i+1}^{U}\alpha^2_{k,g,j}+\sigma_n^2\mathbb{E}[\mathrm{tr}\{[\mathbf{H}^{q\dagger}_{k,g,u}(\mathbf{H}^{q\dagger}_{k,g,u})^H]_{gg}\}]} \\ &= \frac{\frac{1}{\|[\mathbf{H}^{q\dagger}_{k,g,u}]_{g*}\|^2}\alpha^2_{k,g,i}}{\frac{1}{\|[\mathbf{H}^{q\dagger}_{k,g,u}]_{g*}\|^2}\sum_{j=i+1}^{U}\alpha^2_{k,g,j}+\sigma_n^2}.\end{aligned} \tag{H-1}$$

Since $q \in \{1,\cdots,Q\}$ corresponds to the signal reception with the highest effective channel gain magnitude, we define

$$\bar{\varsigma}_{k,g,u} = \max\left\{\varsigma^{1}_{k,g,u},\cdots,\varsigma^{Q}_{k,g,u}\right\}. \tag{H-2}$$

where $\varsigma^{q}_{k,g,u} = \frac{1}{\|[\mathbf{H}^{q\dagger}_{k,g,u}]_{g*}\|^2}$, for $1 \leq q \leq Q$. Now, replacing (H-2) in (H-1) and denoting the transmit SNR by $\rho = \frac{1}{\sigma_n^2}$, we obtain

$$\gamma^{i}_{k,g,u} = \frac{\bar{\varsigma}_{k,g,u}\alpha^2_{k,g,i}}{\bar{\varsigma}_{k,g,u}\sum_{j=i+1}^{U}\alpha^2_{k,g,j}+\frac{1}{\rho}}. \tag{H-3}$$

Note that, since the user U is the strongest one, when $i = u = U$, the i-th message will be recovered without any interference. Then, we can represent the term corresponding to the power of interfering users in (H-3) as

$$\mathscr{I}_i = \begin{cases}\sum_{j=i+1}^{U}\alpha^2_{k,g,j}, & \text{for} \quad 1 \leq i \leq u < U, \\ 0, & \text{for} \quad i = u = U.\end{cases} \tag{H-4}$$

Finally, by substituting (H-4) in (H-3), the SINR expression can be attained as

$$\gamma^{i}_{k,g,u} = \frac{\bar{\varsigma}_{k,g,u}\alpha^2_{k,g,i}}{\bar{\varsigma}_{k,g,u}\mathscr{I}_i+\frac{1}{\rho}}, \qquad 1 \leq i \leq u \leq U, \tag{H-5}$$

which completes the proof.

APPENDIX I – PROOF OF PROPOSITION 4.1

By replacing (53) in (55) and performing some algebraic manipulations, we get the following

$$P_{k,g,u}^{\text{out}} = \Pr\left[\log_2\left(1+\frac{\bar{\varsigma}_{k,g,u}\alpha_{k,g,i}^2}{\bar{\varsigma}_{k,g,u}\mathscr{I}_i+\frac{1}{\rho}}\right) < \mathscr{R}_{k,g,i}\right]$$
$$= \Pr\left[\bar{\varsigma}_{k,g,u} < \frac{2^{\mathscr{R}_{k,g,i}}-1}{\rho[\alpha_{k,g,i}^2-\mathscr{I}_i(2^{\mathscr{R}_{k,g,i}}-1)]}\right] = P\left[\bar{\varsigma}_{k,g,u} < \rho^{-1}\Upsilon_{k,g,u}\right], \tag{I-1}$$

where

$$\Upsilon_{k,g,u} = \max_{1\leq i\leq u}\left\{\frac{2^{\mathscr{R}_{k,g,i}}-1}{\alpha_{k,g,i}^2-\mathscr{I}_i(2^{\mathscr{R}_{k,g,i}}-1)}\right\}. \tag{I-2}$$

The expression (I-1) corresponds to the cumulative distribution function (CDF) of $\bar{\varsigma}_{k,g,u}$. By analyzing (H-1), one can verify that the effective channel gain obtained at each reception q, for $q = 1,\cdots,Q$, is equivalent to the inverse of the g-th main diagonal element of the matrix $\mathbf{\Omega}_{k,g,u} = \mathbf{H}_{k,g,u}^{q\dagger}(\mathbf{H}_{k,g,u}^{q\dagger})^H \in \mathbb{C}^{N\times N}$, which can be expanded as

$$\mathbf{\Omega}_{k,g,u} = (\mathbf{B}_k^H\mathbf{H}_{k,g,u}^q(\mathbf{H}_{k,g,u}^q)^H\mathbf{B}_k)^{-1}\mathbf{B}_k^H\mathbf{H}_{k,g,u}^q(\mathbf{H}_{k,g,u}^q)^H\mathbf{B}_k(\mathbf{B}_k^H\mathbf{H}_{k,g,u}^q(\mathbf{H}_{k,g,u}^q)^H\mathbf{B}_k)^{-1}$$
$$= (\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}. \tag{I-3}$$

As stated in [7], the matrix in (I-3) follows the inverse Wishart distribution and, consequently, the inverse of its main diagonal elements follows the Gamma distribution [82]. As a result, the effective channel gains delivered by the Q sub-arrays can be seen as Q independent and identically distributed Gamma random variables. Therefore, first considering unordered gains, the CDF of $\max\left\{\varsigma_{k,g,u}^1,\cdots,\varsigma_{k,g,u}^Q\right\}$ is given by

$$F_{\max}(x) = \left[\frac{\gamma\left(N-\bar{M}+1, x[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg}\right)}{\Gamma(N-\bar{M}+1)}\right]^Q, \tag{I-4}$$

and its probability density function (PDF) can be derived as

$$f_{\max}(x) = \frac{Q[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg}^{N-\bar{M}+1}}{\Gamma(N-\bar{M}+1)^Q}x^{N-\bar{M}}e^{-x[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg}}$$
$$\times\,\gamma(N-\bar{M}+1, x[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg})^{Q-1}. \tag{I-5}$$

Consequently, the PDF for the ordered effective channel gains $\bar{\varsigma}_{k,g,u}$ can be obtained as

$$\begin{aligned} f_{\bar{\varsigma}_{k,g,u}}(x) &= \sum_{j=0}^{U-u} U \binom{U-1}{u-1} \binom{U-u}{j} (-1)^j f_{\max}(x) F_{\max}(x)^{u-1+j} \\ &= \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} (-1)^j \frac{Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}^{N-\bar{M}+1}}{\Gamma(N-\bar{M}+1)^{Q(u+j)}} x^{N-\bar{M}} e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \\ &\times \gamma(N-\bar{M}+1, x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^{Q(u+j)-1}, \end{aligned} \tag{I-6}$$

where, for easy of notation, we have defined $\mathscr{U}_u = U \binom{U-1}{u-1}$.

At last, the closed-form expression for the general outage probability of the proposed system is obtained by integrating (I-6), as follows

$$\begin{aligned} P_{k,g,u}^{\text{out}} &= \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} (-1)^j \frac{Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}^{N-\bar{M}+1}}{\Gamma(N-\bar{M}+1)^{Q(u+j)}} \\ &\times \int_0^{\rho^{-1}\Upsilon_{k,g,u}} x^{N-\bar{M}} e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \gamma(N-\bar{M}+1, x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^{Q(u+j)-1} dx \\ &= \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} \frac{(-1)^j}{u+j} \left[\frac{\gamma\left(N-\bar{M}+1, \rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}\right)}{\Gamma(N-\bar{M}+1)} \right]^{Q(u+j)}, \end{aligned} \tag{I-7}$$

which completes the proof.

APPENDIX J – PROOF OF PROPOSITION 4.2

By applying the series representation of the Gamma function in (56), we have [86]

$$\begin{aligned} P_{k,g,u}^{\text{out}} &= \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} \frac{(-1)^j}{(u+j)[(N-\bar{M})!]^{Q(u+j)}} [(N-\bar{M})!]^{Q(u+j)} \\ &\times \left(1 - e^{-\rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \sum_{n=0}^{N-\bar{M}} \frac{(\rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^n}{n!}\right)^{Q(u+j)} \\ &= \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} \frac{(-1)^j}{u+j} \left(e^{-\rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}\right. \\ &\times \left. \sum_{n=N-\bar{M}+1}^{\infty} \frac{(\rho^{-1}\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^n}{n!}\right)^{Q(u+j)}. \end{aligned} \tag{J-1}$$

Then, by exploring properties of the Taylor's series and performing some algebraic manipulations in (J-1), we obtain the desired high-SNR approximation as

$$P_{k,g,u}^{\text{out}} \approx \frac{\mathscr{U}_u}{u} \frac{[\Upsilon_{k,g,u}[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}]^{(N-\bar{M}+1)Qu}}{\rho^{(N-\bar{M}+1)Qu}[(N-\bar{M}+1)!]^{Qu}}. \tag{J-2}$$

Consequently, the diversity order experienced by the u-th user is determined by

$$\mathscr{D}_u = (N-\bar{M}+1)\,Qu, \tag{J-3}$$

which completes the proof.

APPENDIX K – PROOF OF PROPOSITION 4.3

The expression in (59) can be rearranged as

$$\bar{C}_{k,g} = \mathbb{E}\left[\sum_{u=1}^{U} \log_2\left(\frac{1+\bar{\varsigma}_{k,g,u}\rho(\alpha_{k,g,u}^2+\mathscr{I}_u)}{1+\bar{\varsigma}_{k,g,u}\rho\mathscr{I}_u}\right)\right]. \tag{K-1}$$

By invoking the PDF of ordered channel gains in (I-6), and defining $\varepsilon_{1,u} = \rho(\alpha_{k,g,u}^2+\mathscr{I}_u)$ and $\varepsilon_{2,u} = \rho\mathscr{I}_u$, the exact ergodic sum-rate can be evaluated by

$$\begin{aligned}\bar{C}_{k,g} &= \sum_{u=1}^{U}\mathbb{E}\left[\log_2\left(\frac{1+\bar{\varsigma}_{k,g,u}\varepsilon_{1,u}}{1+\bar{\varsigma}_{k,g,u}\varepsilon_{2,u}}\right)\right] \\ &= \sum_{u=1}^{U}\mathscr{U}_u\sum_{j=0}^{U-u}\binom{U-u}{j}(-1)^j\frac{Q[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg}^{N-\bar{M}+1}}{\Gamma(N-\bar{M}+1)^{Q(u+j)}}\int_0^{\infty}\log_2\left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}}\right) \\ &\times x^{N-\bar{M}}e^{-x[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg}}\gamma(N-\bar{M}+1,x[(\mathbf{B}_k^H\mathbf{R}_k\mathbf{B}_k)^{-1}]_{gg})^{Q(u+j)-1}dx. \end{aligned}\tag{K-2}$$

which completes the proof.

APPENDIX L – PROOF OF PROPOSITION 4.4

When $N = \bar{M}$, the PDF for the effective channel gains in (I-6) can be simplified to

$$f_{\bar{\varsigma}_{k,g,u}}(x) = \mathscr{U}_u \sum_{j=0}^{U-u} \binom{U-u}{j} (-1)^j Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg} e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \times \gamma(1, x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg})^{Q(u+j)-1}. \quad \text{(L-1)}$$

Then, by expanding the incomplete gamma function in (L-1) by its series representation [86] and performing some manipulations, the PDF becomes

$$\begin{aligned} f_{\bar{\varsigma}_{k,g,u}}(x) &= \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} (-1)^j Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg} e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} \\ &\quad \times (-e^{-x[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}})^i \binom{Q(u+j)-1}{i} \\ &= \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} (-1)^j Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg} \binom{Q(u+j)-1}{i} (-1)^i \\ &\quad \times e^{-x(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}. \end{aligned} \quad \text{(L-2)}$$

Given the above simplification, the ergodic sum-rate expression obtained in (K-2) can be rewritten as

$$\begin{aligned} \bar{C}_{k,g} &= \sum_{u=1}^{U} \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} (-1)^j Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg} \binom{Q(u+j)-1}{i} (-1)^i \\ &\quad \times \int_0^\infty \log_2 \left(\frac{1+x\varepsilon_{1,u}}{1+x\varepsilon_{2,u}} \right) e^{-x(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} dx \\ &= \sum_{u=1}^{U} \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} \frac{(-1)^j Q[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\ln(2)} \binom{Q(u+j)-1}{i} (-1)^i \\ &\quad \times \left[\int_0^\infty \ln(1+x\varepsilon_{1,u}) e^{-x(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} dx - \int_0^\infty \ln(1+x\varepsilon_{2,u}) e^{-x(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}} dx \right]. \end{aligned} \quad \text{(L-3)}$$

The integrals in (L-3) are of the form $\int_0^\infty \ln(1+\beta x) e^{-\mu x} dx$, which solution is given in [86]. Then, by applying the refereed result, we can finally obtain the desired expression for

the ergodic sum-rate, as follows

$$\bar{C}_{k,g} = \sum_{u=1}^{U} \mathscr{U}_u \sum_{j=0}^{U-u} \sum_{i=0}^{Q(u+j)-1} \binom{U-u}{j} \frac{(-1)^j Q}{(i+1)\ln(2)} \binom{Q(u+j)-1}{i} (-1)^i$$
$$\times \left[\mathrm{Ei}\left(-\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{2,u}} \right) e^{\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{2,u}}} \right.$$
$$\left. -\mathrm{Ei}\left(-\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{1,u}} \right) e^{\frac{(i+1)[(\mathbf{B}_k^H \mathbf{R}_k \mathbf{B}_k)^{-1}]_{gg}}{\varepsilon_{1,u}}} \right], \tag{L-4}$$

which completes the proof.