

Massive MIMO-NOMA Networks with Multi-Polarized Antennas

Sousa de Sena Arthur, Benevides da Costa Daniel, Ding Zhiguo, Nardelli Pedro H.
J.

This is a Final draft version of a publication
published by IEEE
in IEEE Transactions on Wireless Communications

DOI: 10.1109/TWC.2019.2937868

Copyright of the original publication: © 2019 IEEE

Please cite the publication as follows:

Sousa de Sena, A., Benevides da Costa, D., Ding, Z., Nardelli, P.H.J. (2019). Massive MIMO-NOMA Networks with Multi-Polarized Antennas. IEEE Transactions on Wireless Communications. DOI: 10.1109/TWC.2019.2937868

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**This is a parallel published version of an original publication.
This version can differ from the original published article.**

Massive MIMO-NOMA Networks with Multi-Polarized Antennas

Arthur Sousa de Sena, *Student Member, IEEE*, Daniel Benevides da Costa,
Senior Member, IEEE, Zhiguo Ding, *Senior Member, IEEE*, and Pedro H. J.
Nardelli, *Member, IEEE*

Abstract

This paper aims to design and evaluate the performance of multi-cluster multi-user dual-polarized massive multiple-input multiple-output (MIMO) systems with non-orthogonal multiple access (NOMA). Assuming the downlink mode in which a single base station communicates with multiple users, with all terminals being equipped with multiple co-located dual-polarized antennas, 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. 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.

Index Terms

Non-orthogonal multiple access (NOMA), massive MIMO, multi-polarized antennas.

A. S. de Sena and D. B. da Costa are with the Federal University of Ceará, Brazil (email: arthurss@alu.ufc.br, danielb-costa@ieee.org).

Z. Ding is with the University of Manchester, UK (email: zhiguo.ding@manchester.ac.uk).

P. H. J. Nardelli is with the Lappeenranta University of Technology, Finland (email: pedro.nardelli@lut.fi).

I. INTRODUCTION

An unprecedented explosion in the number of mobile connections is expected in the upcoming years [1]. Together with this massive number of connected devices, new services and applications, such as virtual reality, ultra-high definition video, and self-driving cars, will require extremely high data-rates, ultra-low latency, and high reliability [2]–[4]. In order to attend these demanding requirements, 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]–[8]. Briefly speaking, a base station (BS) superposes the messages of each user through superposition coding 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) and, more recently, in the 3GPP Release-14, where fifteen NOMA uplink schemes were proposed [5], [6].

Massive multiple-input multiple-output (MIMO) is another enabling technology for 5G that can boost the system capacity by multiplexing multiple users through spatial domain. In combination with NOMA, massive MIMO systems can provide even higher spectral improvements that can remarkably outperform traditional orthogonal multiple access (OMA) protocols [7], [8]. The performance gains achieved in MIMO systems are scaled with the increase in the number of antennas. However, if the antenna elements are not separated by at least half of the wavelength, the system performance is severely degraded [9]–[11]. Due to this impairing characteristic, in practical massive MIMO implementations, the maximum number of antennas is limited. To alleviate this issue, several works have considered the installation of co-located orthogonal polarized antennas [9]–[13]. Such strategy enables the design of compact massive arrays with very low correlation between orthogonal antenna elements. 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 [9], [12], [13]. 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 [14]–[16].

A. Related Works

Considering a single cell deployment, Ding *et al.* [7] employed NOMA to a downlink multi-user MIMO scenario assuming fewer antennas at the BS than at the users. On the other hand, in [8] a multi-user massive MIMO-NOMA scenario was investigated assuming that the number of transmit antennas was higher than the ones at the users. In addition, 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 [17], 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 [18]–[20]. A new non-regenerative massive MIMO-NOMA relay system design was proposed in [18], 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 [19], the authors proposed a suboptimal algorithm to maximize the cell-edge users' achievable rate, and in [20] the performance of a multi-relay massive MIMO-NOMA system was analyzed. The single-input single-output (SISO) NOMA case was considered in [21] and [22]. In [21], by grouping users into multiple pairs, the authors have derived optimum power allocation aiming various objectives, such as fairness maximization, sum-rate maximization, and energy efficiency maximization, and in [22], for a single multi-user NOMA group, the optimum power allocation for weighted sum-rate maximization was obtained.

The design of polarized schemes has gained increasing attention over the last years. In [9], 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 [12], [13], 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 [10], [16], [23], [24]. The employment of polarized antennas in MIMO-NOMA was only considered in [25]. 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 our work. Besides, only simulation results were presented, lacking analytical derivation.

B. Motivation and Contributions

Although there is an uncountable number of relevant works for MIMO-NOMA and multi-polarized MIMO separately, to the best of the authors' knowledge, a full and in-depth understanding of the combination of these two subjects has not been examined in the literature so far. Therefore, motivated by this lacuna, this paper aims to design and analyze the performance of multi-cluster multi-user dual-polarized massive MIMO-NOMA systems. Specifically, 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, from which we obtain a novel analytical framework that provides a simpler and more practical way than system simulation 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.

Notation and Special Functions: Bold-faced lower-case letters represent vectors and upper-case letters denote matrices. The norm and the i -th element of a vector \mathbf{a} are represented by $\|\mathbf{a}\|$ and $[\mathbf{a}]_i$, respectively. The notations $[\mathbf{A}]_{i,j}$ and $[\mathbf{A}]_{i,*}$ correspond the (i, j) entry and the i -th row of the matrix \mathbf{A} , respectively. The Hermitian transposition of a matrix \mathbf{A} is denoted by \mathbf{A}^H and the trace by $\text{tr}\{\mathbf{A}\}$. \mathbf{I}_M represents the identity matrix of dimension $M \times M$. In addition, \otimes represents the Kronecker product, $E[\cdot]$ denotes **expectation**, $\Gamma(\cdot)$ is the Gamma function, and $\gamma(\cdot, \cdot)$ corresponds to the lower incomplete Gamma function.

II. SYSTEM MODEL

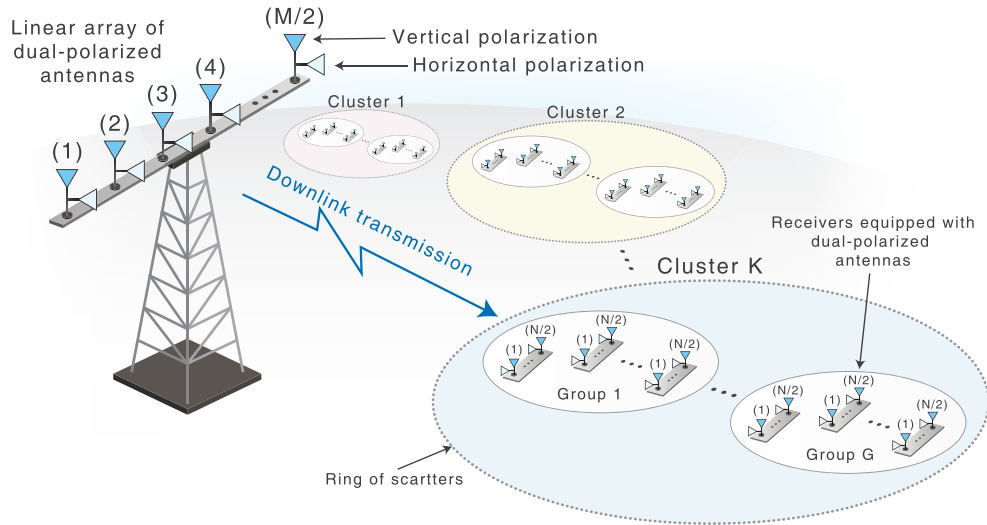


Fig. 1: System model. Both BS and users are equipped with multiple co-located dual-polarized antennas.

Let a downlink transmit scenario in which a single BS communicates with multiple users, as illustrated in Fig. 1. 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 [8], [9], [26]. 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 [9], [10]

$$\mathbf{H}_{k,l} = \begin{bmatrix} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{v \rightarrow v} & \sqrt{\chi} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{h \rightarrow v} \\ \sqrt{\chi} \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{v \rightarrow h} & \mathbf{U}_k \mathbf{\Lambda}_k^{\frac{1}{2}} \mathbf{G}_{k,l}^{h \rightarrow h} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{k,l}^{v \rightarrow v} & \mathbf{H}_{k,l}^{h \rightarrow v} \\ \mathbf{H}_{k,l}^{v \rightarrow h} & \mathbf{H}_{k,l}^{h \rightarrow h} \end{bmatrix} \in \mathbb{C}^{M \times N}, \quad (1)$$

with covariance matrix given by

$$\mathbf{R}_k = (\chi + 1) \begin{bmatrix} \tilde{\mathbf{R}}_k & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{R}}_k \end{bmatrix}, \quad (2)$$

where $\chi \in \{0, 1\}$ stands for the inverse of the antennas' cross-polar discrimination parameter (XPD), $\tilde{\mathbf{R}}_k \in \mathbb{C}^{\frac{M}{2} \times \frac{M}{2}}$ represents the covariance matrix of rank r_k , corresponding to one polarization, $\mathbf{\Lambda}_k$ is an $r_k^* \times r_k^*$ diagonal matrix containing r_k^* nonzero eigenvalues of $\tilde{\mathbf{R}}_k$, with elements ordered in decreasing order, $\mathbf{U}_k \in \mathbb{C}^{\frac{M}{2} \times r_k^*}$ denotes a matrix composed of eigenvectors of $\tilde{\mathbf{R}}_k$ corresponding to the nonzero eigenvalues, and $\mathbf{G}_{k,l}^{i \rightarrow j} \in \mathbb{C}^{r_k^* \times \frac{N}{2}}$ is the channel matrix of the link between the i -th polarized transmitting antennas and the j -th polarized receiving antennas, $i, j \in \{v, h\}$, 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, similarly to the strategy presented in [8], 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

$$\mathbf{x} = \sum_{k=1}^K \mathbf{B}_k \sum_{g=1}^G \mathbf{p}_{k,g} \sum_{u=1}^U \alpha_{k,g,u} s_{k,g,u}, \quad (3)$$

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 state information, e.g., the 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 a precoding vector for the g -th group.

III. PRECODER DESIGNS AND SIGNAL RECEPTION

In order to reduce the feedback overhead and, consequently, alleviate the computational complexity inherent to massive MIMO systems, a two-stage precoder $\mathbf{B}_k \mathbf{p}_{k,g}$ is proposed. The outer preprocessing matrix \mathbf{B}_k has the role of eliminating the interferences from other clusters, while the inner precoding vector $\mathbf{p}_{k,g}$ is intended 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

$$\mathbf{H}_{k,l}^H \mathbf{B}_{k'} \approx \mathbf{0}, \forall k \neq k'. \quad (4)$$

To accomplish this goal, the subspace spanned by the preprocessing matrix must be near-orthogonal (perfect orthogonality is achieved when $r_k^* = r_k$) to the subspace spanned by the dominant eigenvectors of other clusters [27], 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 $r_1^* = r_2^* = \dots = r_K^* = r^*$, $r_1 = r_2 = \dots = r_K = r$ and we define the following matrix

$$\mathbf{U}_k^- = [\mathbf{U}_1, \dots, \mathbf{U}_{k-1}, \mathbf{U}_{k+1}, \dots, \mathbf{U}_K] \in \mathbb{C}^{\frac{M}{2} \times (K-1)r^*}. \quad (5)$$

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 $\mathbf{E}_k^{(0)} \in \mathbb{C}^{\frac{M}{2} \times \frac{M}{2} - (K-1)r^*}$ represents the eigenvectors that correspond to the last $\frac{M}{2} - (K-1)r^*$ 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^* eigenmodes of the groups $k' \neq k$ can be defined as $\hat{\mathbf{H}}_{k,l} = (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)})^H \mathbf{H}_{k,l}$, where \mathbf{I}_2 is a 2×2 identity matrix. Consequently, the covariance

matrix of $\hat{\mathbf{H}}_{k,l}$ is given by

$$\hat{\mathbf{R}}_k = (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)})^H \mathbf{R}_k (\mathbf{I}_2 \otimes \mathbf{E}_k^{(0)}), \quad (6)$$

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

$$\underline{\hat{\mathbf{R}}}_k = (\mathbf{E}_k^{(0)})^H \tilde{\mathbf{R}}_k \mathbf{E}_k^{(0)}. \quad (7)$$

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 $\mathbf{F}_k^{(1)} \in \mathbb{C}^{\frac{M}{2} - (K-1)r^* \times \frac{\bar{M}}{2}}$ consists of the first $\frac{\bar{M}}{2}$ columns of \mathbf{F}_k , the precoder design can be obtained as follows

$$\tilde{\mathbf{B}}_k = \mathbf{E}_k^{(0)} \mathbf{F}_k^{(1)} \in \mathbb{C}^{\frac{M}{2} \times \frac{\bar{M}}{2}}, \quad (8)$$

$$\mathbf{B}_k = \begin{bmatrix} \tilde{\mathbf{B}}_k & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{B}}_k \end{bmatrix} \in \mathbb{C}^{M \times \bar{M}}, \quad (9)$$

in which the following constraints must hold

$$K \leq \bar{M} \leq 2 \left(\frac{M}{2} - (K-1)r^* \right), \quad (10)$$

and

$$\bar{M} \leq 2r^* \leq 2r. \quad (11)$$

It is important to highlight that r^* 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^* gets to r , less inter-cluster interference the system will experience. However, as stated in [27], choosing r^* too large, e.g., $r^* = r$, does not bring significant improvements. Instead, it will only increase the factor $(K-1)r^*$ and, since the constraint $(K-1)r^* < \frac{M}{2}$ must be obeyed, this will eventually force a reduction on the number of clusters that are simultaneously served. Therefore, the parameter r^* should be carefully chosen in order to attend the system requirements. In our implementation, given the constraints in (10) and (11), we choose to configure the dominant eigenvalues as $r^* = \min \left\{ r, \left\lfloor \left(\frac{M}{2} - \frac{\bar{M}}{2} \right) \frac{1}{K-1} \right\rfloor \right\}$.

Since it is not an easy task to acquire the fast-fading matrices at the transmitter and in order

to further 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.

A. 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 antenna element 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

$$\mathbf{p}_{k,g} = [0, \dots, \underset{[\mathbf{p}_{k,g}]_{g-1}}{\uparrow} 0, \underset{[\mathbf{p}_{k,g}]_g}{\uparrow} 1, \underset{[\mathbf{p}_{k,g}]_{g+1}}{\uparrow} 0, \dots, 0]^T. \quad (12)$$

Note that, with the above precoder¹, the g -th **effective data stream** is assigned to the g -th group.

B. 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

$$\mathbf{p}_{k,g} = \begin{bmatrix} \mathbf{p}_{k,g}^v \\ \mathbf{p}_{k,g}^h \end{bmatrix} = \begin{bmatrix} [0, \dots, \underset{[\mathbf{p}_{k,g}]_{g-1}}{\uparrow} 0, \underset{[\mathbf{p}_{k,g}]_g}{\uparrow} 1, \underset{[\mathbf{p}_{k,g}]_{g+1}}{\uparrow} 0, \dots, 0]^T \\ [0, \dots, \underset{[\mathbf{p}_{k,g}]_{g-1}}{\uparrow} 0, \underset{[\mathbf{p}_{k,g}]_g}{\uparrow} 1, \underset{[\mathbf{p}_{k,g}]_{g+1}}{\uparrow} 0, \dots, 0]^T \end{bmatrix}, \quad (13)$$

where $\mathbf{p}_{k,g}^p \in \mathbb{C}^{\frac{\bar{M}}{2} \times 1}$ 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.

¹In order to provide insightful performance comparisons, similarly as in a conventional 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 III.

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 to the user's terminal.

C. 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

$$\mathbf{y}_{k,g,u} = \begin{bmatrix} \mathbf{y}_{k,g,u}^v \\ \mathbf{y}_{k,g,u}^h \end{bmatrix} = \mathbf{H}_{k,g,u}^H \sum_{l=1}^K \mathbf{B}_l \sum_{m=1}^G \mathbf{p}_{l,m} \sum_{n=1}^U \alpha_{l,m,n} s_{l,m,n} + \mathbf{n}_{k,g,u}, \quad (14)$$

where $\mathbf{y}_{k,g,u}^v \in \mathbb{C}^{\frac{N}{2} \times 1}$ and $\mathbf{y}_{k,g,u}^h \in \mathbb{C}^{\frac{N}{2} \times 1}$ 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^* , the near-orthogonality condition of (4) can be satisfied [9], [27]. 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 (14) as

$$\mathbf{y}_{k,g,u} = \mathbf{H}_{k,g,u}^H \mathbf{B}_k \sum_{m=1}^G \mathbf{p}_{k,m} \sum_{n=1}^U \alpha_{k,m,n} s_{k,m,n} + \mathbf{n}_{k,g,u}. \quad (15)$$

Now, in order to eliminate the inter-group interference and recover the desired signal, a zero-forcing equalizer 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 (15) 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.,

$$\hat{\mathbf{s}}_{k,g,u} = \sum_{m=1}^G \mathbf{p}_{k,m} \sum_{n=1}^U \alpha_{k,m,n} s_{k,m,n} + \bar{\mathbf{H}}_{k,g,u}^\dagger \mathbf{n}_{k,g,u}. \quad (16)$$

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 (12) or as (13), and this is further detailed next.

Approach I: It is considered that the system is operating on multiplexing mode, that is, the precoder in (12) 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 (16) can be rewritten as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \sum_{n=1}^U \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^U \alpha_{k,G,n} s_{k,G,n} \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^\dagger \mathbf{n}_{k,g,u}. \quad (17)$$

One can see that each element of (17) 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

$$\hat{s}_{k,g,u} = [\hat{\mathbf{s}}_{k,g,u}]_g = \sum_{n=1}^U \alpha_{k,g,n} s_{k,g,n} + [\bar{\mathbf{H}}_{k,g,u}^\dagger \mathbf{n}_{k,g,u}]_g. \quad (18)$$

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 (13), 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 (16) can be expressed as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \sum_{n=1}^U \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^U \alpha_{k,G,n} s_{k,G,n} \\ \sum_{n=1}^U \alpha_{k,1,n} s_{k,1,n} \\ \vdots \\ \sum_{n=1}^U \alpha_{k,G,n} s_{k,G,n} \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^\dagger \mathbf{n}_{k,g,u} = \begin{bmatrix} \mathbf{s}_k^v \\ \mathbf{s}_k^h \end{bmatrix} + \bar{\mathbf{H}}_{k,g,u}^\dagger \mathbf{n}_{k,g,u}, \quad (19)$$

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 (1) 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

$$\bar{\mathbf{H}}_{k,g,u} = \begin{bmatrix} (\mathbf{H}_{k,g,u}^{v \rightarrow v})^H \tilde{\mathbf{B}}_k & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}_{k,g,u}^{h \rightarrow h})^H \tilde{\mathbf{B}}_k \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^v & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^h \end{bmatrix}. \quad (20)$$

Consequently, its corresponding zero-forcing detection matrix becomes

$$\bar{\mathbf{H}}_{k,g,u}^\dagger = \begin{bmatrix} ((\bar{\mathbf{H}}_{k,g,u}^v)^H \bar{\mathbf{H}}_{k,g,u}^v)^{-1} (\bar{\mathbf{H}}_{k,g,u}^v)^H & \mathbf{0} \\ \mathbf{0} & ((\bar{\mathbf{H}}_{k,g,u}^h)^H \bar{\mathbf{H}}_{k,g,u}^h)^{-1} (\bar{\mathbf{H}}_{k,g,u}^h)^H \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \end{bmatrix}. \quad (21)$$

From above, the signal in (19) can be rewritten as

$$\hat{\mathbf{s}}_{k,g,u} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{k,g,u}^v \\ \mathbf{y}_{k,g,u}^h \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{H}}_{k,g,u}^{\dagger v} \mathbf{y}_{k,g,u}^v \\ \bar{\mathbf{H}}_{k,g,u}^{\dagger h} \mathbf{y}_{k,g,u}^h \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{s}}_{k,g,u}^v \\ \hat{\mathbf{s}}_{k,g,u}^h \end{bmatrix}, \quad (22)$$

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

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

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 antenna selection 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 (22), and this arises as a potential future work.

IV. 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 signal-to-noise ratio (SNR) regime is carried out. Finally, the ergodic sum-rate capacity of both approaches is also investigated.

A. 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 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

$$\gamma_{k,g,u}^i = \frac{\zeta_{k,g,u} \alpha_{k,g,i}^2}{\zeta_{k,g,u} \mathcal{I}_i + \frac{1}{\rho}}, \quad \text{for } 1 \leq i \leq U. \quad (24)$$

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

$$\mathcal{I}_i = \begin{cases} \sum_{j=i+1}^U \alpha_{k,g,j}^2, & \text{for } 1 \leq i \leq u < U, \\ 0, & \text{for } i = u = U, \end{cases} \quad (25)$$

Proof: Please, see Appendix A.

Lemma II (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

$$\gamma_{k,g,u}^i = \frac{\zeta_{k,g,u}^* \alpha_{k,g,i}^2}{\zeta_{k,g,u}^* \mathcal{I}_i + \frac{1}{\rho}}, \quad \text{for } 1 \leq i \leq U, \quad (26)$$

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

Proof: Please, see Appendix B.

One can see that the interference factor defined in (25) reaches its maximum value when the first user, the weakest one, decodes its message. **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.

B. Outage Probability

The outage probability for the u -th user in the g -th group of the k -th cluster, denoted by $P_{k,g,u}^{\text{out}}$, 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 $\mathcal{R}_{k,g,i}$, $\forall 1 \leq i \leq u$, i.e.,

$$P_{k,g,u}^{\text{out}} = P[\log_2(1 + \gamma_{k,g,u}^i) < \mathcal{R}_{k,g,i}], \quad \text{for all } 1 \leq i \leq u. \quad (27)$$

From (27), next closed-form expressions for the outage probability are derived for dual-polarized massive MIMO-NOMA systems assuming the previously discussed approaches.

Proposition I (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

$$P_{k,g,u}^{\text{out}} = \mathcal{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{1}{u+n} \left[\frac{\gamma(\eta, (\chi+1)\beta_{k,g}\rho^{-1}\Upsilon_{k,g,u})}{\Gamma(\eta)} \right]^{u+n}, \quad (28)$$

where $\mathcal{U}_u = U \binom{U-1}{u-1}$, $\eta = \frac{N}{2} - \frac{\bar{M}}{2} + 1$, $\beta_{k,g} = [\mathbf{I}_2 \otimes (\tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k)^{-1}]_{g,g}$ and $\Upsilon_{k,g,u} = \max_{i \in [1,u]} \left\{ \frac{2^{\mathcal{R}_{k,g,i-1}}}{\alpha_{k,g,i}^2 - \mathcal{I}_i(2^{\mathcal{R}_{k,g,i-1}})} \right\}$.

Proof: Please, see Appendix C.

Proposition II (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

$$P_{k,g,u}^{\text{out}} = \mathcal{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{1}{u+n} \left[\frac{\gamma(\eta, (\chi+1)\tilde{\beta}_{k,g}\rho^{-1}\Upsilon_{k,g,u})}{\Gamma(\eta)} \right]^{2(u+n)}, \quad (29)$$

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

Proof: It is similar to the proof of Approach I. Due to space constraints, it is not shown here.

From above, note that the greater exponent in (29) influences the outage probability to decrease faster than in (28), indicating a superior performance of the Approach II.

C. Asymptotic Analysis

In order to gain further insights from the derived expressions, an asymptotic analysis is now performed.

Proposition III:

a) *Asymptotic Outage Probability for Approach I:* When the transmit SNR goes to infinity, i.e., $\rho \rightarrow \infty$, the outage probability expression in (28) can be approximated by

$$P_{k,g,u}^{\text{out}} \approx \frac{1}{\rho^{\eta u}} \frac{\mathcal{U}_u}{u} \frac{[(\chi+1)\beta_{k,g}\Upsilon_{k,g,u}]^{\eta u}}{(\eta!)^u}, \quad (30)$$

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

$$\mathcal{D}_u = \left(\frac{N}{2} - \frac{\bar{M}}{2} + 1 \right) u. \quad (31)$$

b) *Asymptotic Outage Probability for Approach II*: Now, when the transmit SNR goes to infinity, i.e., $\rho \rightarrow \infty$, the outage probability in (29) can be approximated by

$$P_{k,g,u}^{\text{out}} \approx \frac{1}{\rho^{2\eta u}} \frac{\mathcal{U}_u [(\chi + 1)\beta_{k,g} \Upsilon_{k,g,u}]^{2\eta u}}{u (\eta!)^{2u}}, \quad (32)$$

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

$$\mathcal{D}_u = (N - \bar{M} + 2) u. \quad (33)$$

Proof: Please, see Appendix D.

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.

D. 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 $\gamma_{k,g,u}^u$. Therefore, the sum-rate for the g -th group in the k -th cluster can be defined as

$$C_{k,g} = \sum_{u=1}^U \log_2(1 + \gamma_{k,g,u}^u). \quad (34)$$

Proposition IV (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, $\forall i \in \{1, 2, \dots, u\}$, are correctly decoded, the ergodic sum-rate achieved by the g -th group in the k -cluster can be

derived as

$$\begin{aligned} \bar{C}_{k,g} &= \sum_{u=1}^U \mathcal{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{(\chi+1)^\eta \beta_{k,g}^\eta}{\Gamma(\eta)^{u+n}} \\ &\quad \times \int_0^\infty \log_2 \left(\frac{1+x\epsilon_{1,u}}{1+x\epsilon_{2,u}} \right) x^{\eta-1} e^{-(\chi+1)\beta_{k,g}x} \gamma(\eta, (\chi+1)\beta_{k,g}x)^{u-1+n} dx, \end{aligned} \quad (35)$$

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

Proof: Please, see Appendix E.

Proposition V (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

$$\begin{aligned} \bar{C}_{k,g} &= \sum_{u=1}^U \mathcal{U}_u \sum_{n=0}^{U-u} (-1)^n \binom{U-u}{n} \frac{2(\chi+1)^\eta \beta_{k,g}^\eta}{\Gamma(\eta)^{2+2(u-1+n)}} \\ &\quad \times \int_0^\infty \log_2 \left(\frac{1+x\epsilon_{1,u}}{1+x\epsilon_{2,u}} \right) x^{\eta-1} e^{-(\chi+1)\beta_{k,g}x} \gamma(\eta, (\chi+1)\beta_{k,g}x)^{1+2(u-1+n)} dx, \end{aligned} \quad (36)$$

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

Proof: Please, see Appendix F.

E. 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

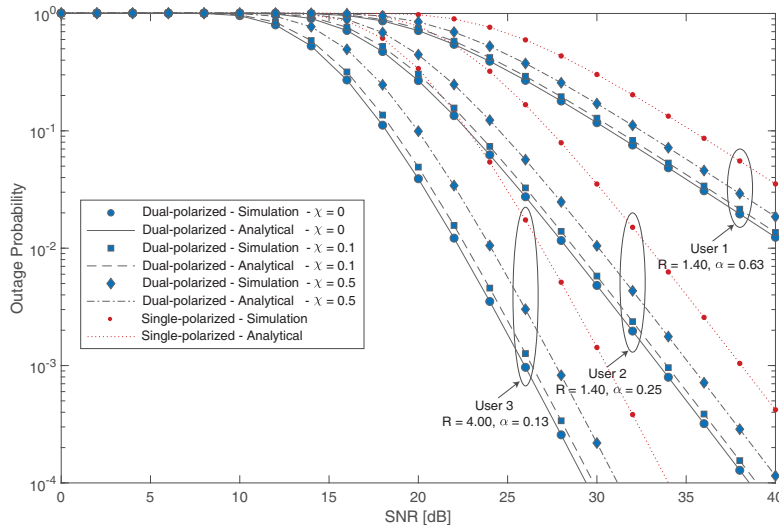


Fig. 2: Outage probability versus transmit SNR for dual-polarized massive MIMO-NOMA system under Approach 1 ($N = 4$).

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.

V. 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 k th cluster is defined as $\theta_k = \frac{\pi}{45} + (\pi - \frac{2\pi}{45}) \frac{k-1}{K-1}$, for $k = 1, \dots, K$, 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 $r^* = 16$. This makes the full matrix in (2) to have $2r^* = 32$ effective eigenvalues, which is the same as that one configured to the single-polarized system. The power allocation coefficients of Users 1, 2 and 3 are set to $\alpha_1^2 = 0.63$, $\alpha_2^2 = 0.25$, $\alpha_3^2 = 0.12$,

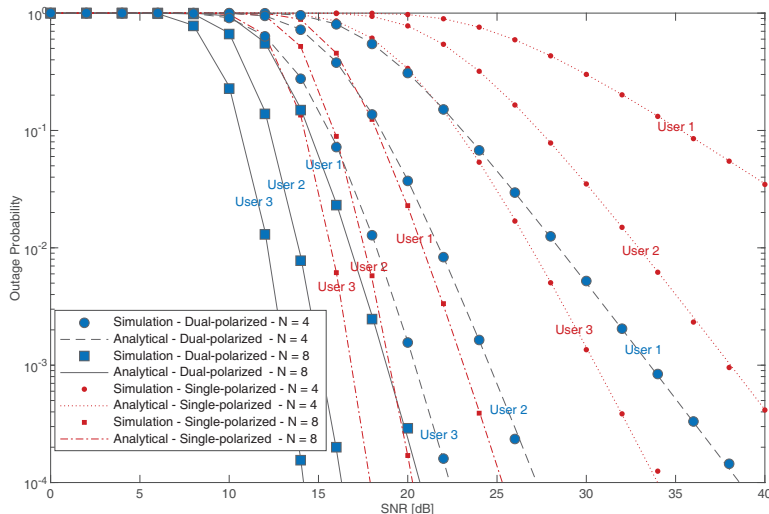


Fig. 3: Outage probability versus transmit SNR for dual-polarized massive MIMO-NOMA system under Approach 2 ($\chi = 0.001$).

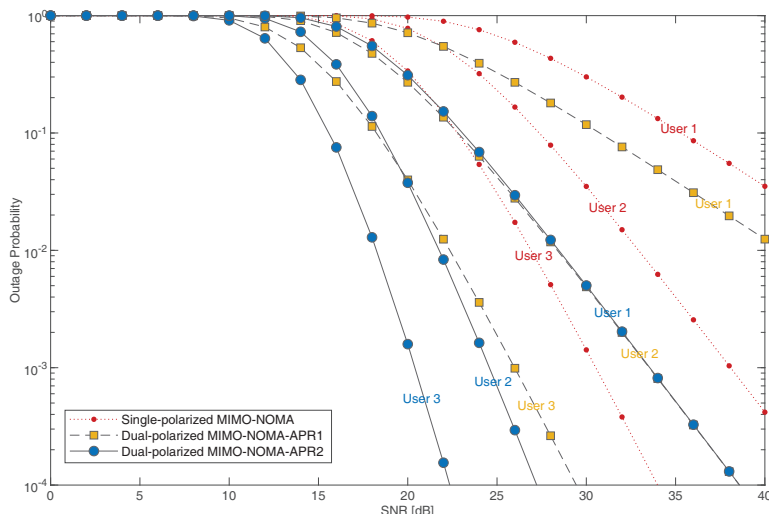


Fig. 4: Outage probability versus transmit SNR for single-polarized and dual-polarized MIMO-NOMA systems under Approaches 1 and 2 ($N = 4$ and $\chi = 0.001$).

and the target rates to $\mathcal{R}_1 = \mathcal{R}_2 = 1.4, \mathcal{R}_3 = 4$ 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.

Fig. 2 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

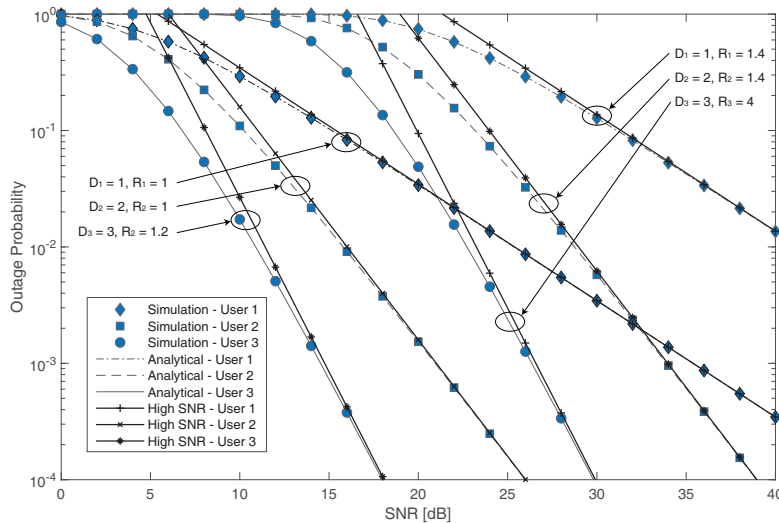


Fig. 5: Exact and asymptotic outage probability curves for dual-polarized massive MIMO-NOMA system under Approach 1 ($N = 4$ and $\chi = 0.1$).

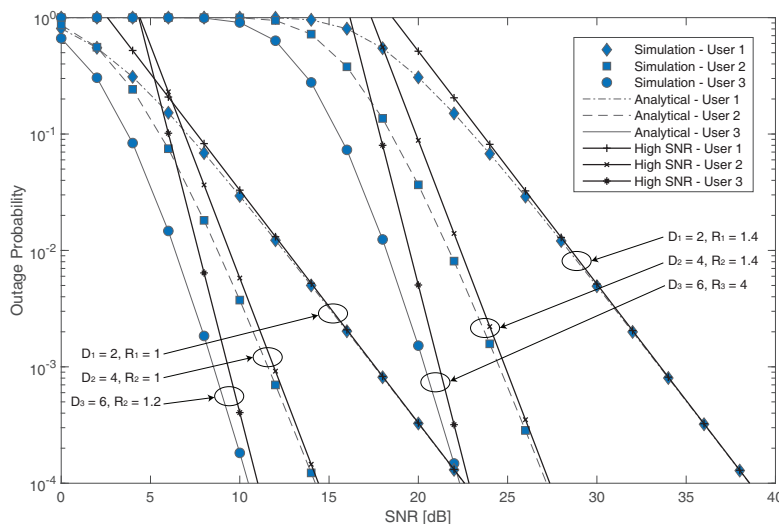


Fig. 6: Exact and asymptotic outage probability curves for dual-polarized massive MIMO-NOMA system under Approach 2 ($N = 4$ and $\chi = 0.1$).

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, Fig. 3 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

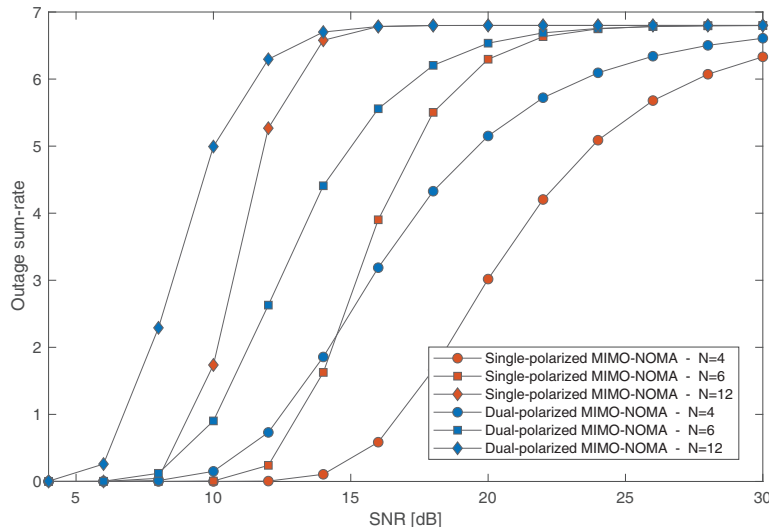


Fig. 7: Outage sum-rate for single-polarized and dual-polarized MIMO-NOMA system under Approach 1 ($\chi = 0.1$).

that this second approach can offer, confirming the insights highlighted in Section IV-B. 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. Fig. 4 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. Also, one can observe that the gain achieved by Approach 1 over the single-polarized case can go up to 4 dB.

Figs. 5 and 6 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 not affect the diversity orders. These behaviors are in complete agreement with the analytical derivation of Proposition III. In addition, as anticipated in Section IV-C, the dual-polarized system under Approach 2 exhibits higher diversity orders than the obtained with approach 1.

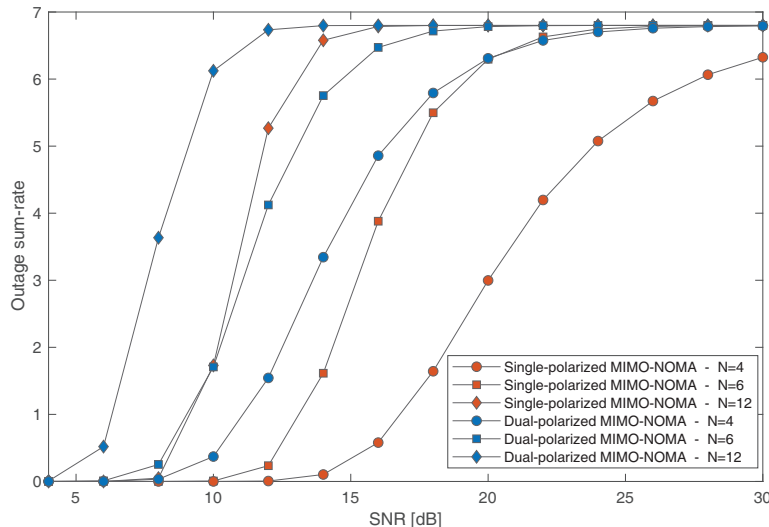


Fig. 8: Outage sum-rate curves for single-polarized and dual-polarized MIMO-NOMA system under Approach 2 ($\chi = 0.001$).

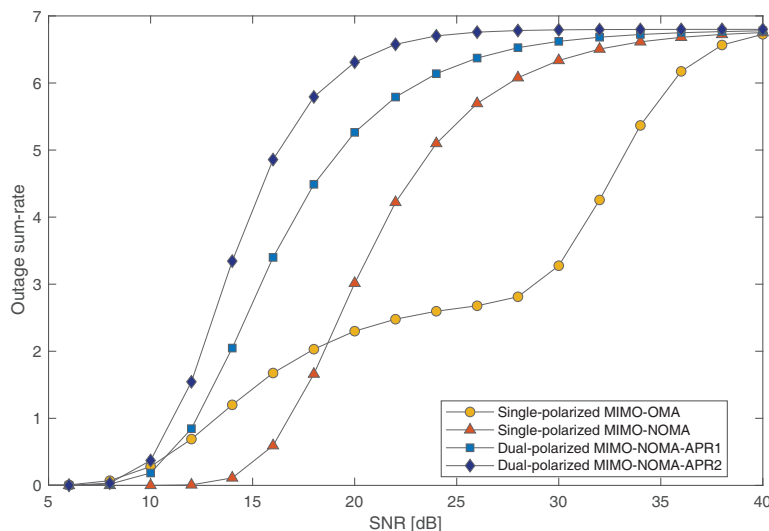


Fig. 9: Outage sum-rate curves for single-polarized MIMO-OMA/MIMO-NOMA systems and dual-polarized MIMO-NOMA systems under Approaches 1 and 2 ($N = 4$ and $\chi = 0.001$).

In Figs. 7 and 8, the outage sum-rate, obtained as $\sum_{n=1}^U (1 - P_{k,g,n}^{\text{out}}) \mathcal{R}_{k,g,n}$, 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 observed, antenna polarization can provide significant spectral improvements, which the gains attained with Approach 2 being even more impressive. For instance, in Fig. 8, 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

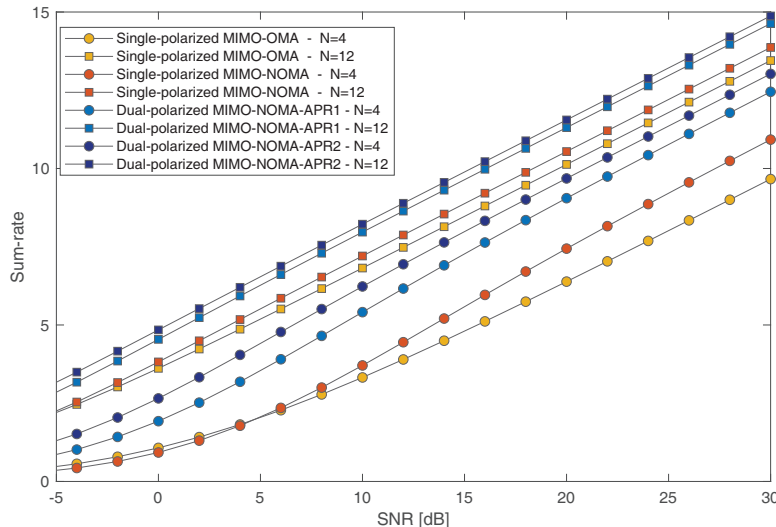


Fig. 10: **Simulated** ergodic sum-rate for various system setups ($\chi = 0.001$).

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 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. Fig. 9 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, Fig. 10 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.

VI. CONCLUSIONS

In this work, we have investigated the application of a dual-polarized antenna array in a massive MIMO-NOMA setup. Two precoder designs were proposed and investigated, and a detailed analytical analysis in terms of outage probability and outage/ergodic sum-rate was provided. 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.

APPENDIX A

PROOF OF LEMMA I

To successfully apply SIC, the effective channel gains of the users within a group must be ordered. Then, assuming that $\zeta_{k,g,1} < \zeta_{k,g,2} < \dots < \zeta_{k,g,U}$, (18) can be rewritten as

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

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

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

Knowing that

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

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

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

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

$$\gamma_{k,g,u}^i = \frac{\frac{1}{\|[\bar{\mathbf{H}}_{k,g,u}^\dagger]_{g,*}\|^2} \alpha_{k,g,i}^2}{\sum_{j=i+1}^U \frac{1}{\|[\bar{\mathbf{H}}_{k,g,u}^\dagger]_{g,*}\|^2} \alpha_{k,g,j}^2 + \frac{1}{\rho}}, \quad \text{for } 1 \leq i \leq u < U. \quad (\text{A-5})$$

From (A-5), the effective channel gain can be defined as $\zeta_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}_{k,g,u}^\dagger]_{g,*}\|^2}$. 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

$$\mathcal{I}_i = \begin{cases} \sum_{j=i+1}^U \alpha_{k,g,j}^2, & \text{for } 1 \leq i \leq u < U, \\ 0, & \text{for } i = u = U. \end{cases} \quad (\text{A-6})$$

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

$$\gamma_{k,g,u}^i = \frac{\zeta_{k,g,u} \alpha_{k,g,i}^2}{\zeta_{k,g,u} \mathcal{I}_i + \frac{1}{\rho}}, \quad \text{for } 1 \leq i \leq U, \quad (\text{A-7})$$

which concludes the proof of Lemma 1.

APPENDIX B

PROOF OF LEMMA II

Similarly to Appendix A, we consider that the effective channel gains are sorted out in increasing order, i.e., $\zeta_{k,g,1}^* < \zeta_{k,g,2}^* < \dots < \zeta_{k,g,U}^*$. Then, (23) can be rearranged as

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

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

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

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

$$\gamma_{k,g,u}^i = \frac{\zeta_{k,g,u}^* \alpha_{k,g,i}^2}{\zeta_{k,g,u}^* \mathcal{I}_i + \frac{1}{\rho}}, \quad \text{for } 1 \leq i \leq U. \quad (\text{B-3})$$

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

APPENDIX C

PROOF OF PROPOSITION I

We first simplify (27) as following

$$P_{k,g,u}^{\text{out}} = P[\log_2(1 + \gamma_{k,g,u}^i) < \mathcal{R}_{k,g,i}] = P[\gamma_{k,g,u}^i < 2^{\mathcal{R}_{k,g,i}} - 1]. \quad (\text{C-1})$$

Then, by replacing (24) in (C-1) and applying simple algebraic manipulations, we get

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

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

$$P_{k,g,u}^{\text{out}} = P[\zeta_{k,g,u} < \rho^{-1} \Upsilon_{k,g,u}]. \quad (\text{C-3})$$

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 $\zeta_{k,g,u} = \frac{1}{\|[\bar{\mathbf{H}}_{k,g,u}^{\dagger}]_{g,*}\|^2}$. This value corresponds to the inverse g -th diagonal element of the covariance matrix $\Xi_{k,g,u} = \bar{\mathbf{H}}_{k,g,u}^{\dagger} \bar{\mathbf{H}}_{k,g,u}^{\dagger H}$. Thus, with the purpose of determining the main diagonal marginal distribution of $\Xi_{k,g,u}$, we perform the following expansion

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

where $\tilde{\mathbf{W}}_k = \tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k$ and $\mathbf{W}_k^{-1} = \mathbf{I}_2 \otimes \tilde{\mathbf{W}}_k^{-1}$. Given that the matrices $\mathbf{G}_{k,l}^{i \rightarrow j}$ have complex Gaussian entries, the resulting matrix $\mathbf{G}_{k,l}^{i \rightarrow j} (\mathbf{G}_{k,l}^{i \rightarrow j})^H$ is complex Wishart distributed. As a consequence, the matrix $\tilde{\mathbf{W}}_k$ also follows the Wishart distribution and, in this way, $\tilde{\mathbf{W}}_k^{-1}$ is inverse Wishart distributed with $\frac{N}{2}$ degrees of freedom and covariance matrix $(\tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k)^{-1}$, that is, $\tilde{\mathbf{W}}_k^{-1} \sim \mathcal{W}_{\frac{M}{2}}^{-1}(\frac{N}{2}, (\tilde{\mathbf{B}}_k^H \tilde{\mathbf{R}}_k \tilde{\mathbf{B}}_k)^{-1})$ [8], [28].

At this point, we already know the distribution of $\tilde{\mathbf{W}}_k^{-1}$. However, we are interested in the distribution of the inverse of the main diagonal elements of $\Xi_{k,g,u}$. 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 [28], [29]. Thus, the inverse of the diagonal elements of \mathbf{W}_k^{-1} follows the Gamma distribution. Consequently, considering the unordered case, the CDF for the inverse of diagonal elements of $\Xi_{k,g,u}$ is given by [30]

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

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

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

where $\eta = \frac{N}{2} - \frac{M}{2} + 1$ 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 [28]

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

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 $\mathcal{U}_u = U \binom{U-1}{u-1}$, (C-7) can be simplified as

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

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 be derived as

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

which completes the proof.

APPENDIX D

PROOF OF PROPOSITION III

Since $\eta = \frac{N}{2} - \frac{\bar{M}}{2} + 1$, and considering that $\frac{N}{2} \geq \frac{\bar{M}}{2}$, η assumes only positive integer values. Then, a series representation for this special case can be applied to simplify the Gamma functions presented in (28) and (29), as follows [31], [32]

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

Thus, the high-SNR outage probability approximations for Approaches I and II can be obtained by replacing (D-1) in (28) and (29), which results in

a) Asymptotic Outage Probability for Approach I:

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

The diversity order reached by the u -th user is

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

b) *Asymptotic Outage Probability for Approach II:*

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

The diversity order reached by the u -th user is

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

APPENDIX E

PROOF OF PROPOSITION IV

By replacing (24) in (34) and performing some simplifications, we have

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

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

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

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

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

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

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

which completes the proof.

APPENDIX F

PROOF OF PROPOSITION V

Similarly to the analysis employed to Approach I, the sum-rate expression in (34) can be simplified for Approach 2 as

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

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

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

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

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

which completes the proof.

REFERENCES

- [1] J. Stryjak and M. Sivakumaran, "The mobile economy 2019." *GSMA Intelligence*, Feb. 2019.
- [2] A. de la Fuente, R. P. Leal, and A. G. Armada, "New technologies and trends for next generation mobile broadcasting services," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 217–223, Nov. 2016.
- [3] G. Liu, Y. Huang, F. Wang, J. Liu, and Q. Wang, "5G features from operation perspective and fundamental performance validation by field trial," *China Commun.*, vol. 15, no. 11, pp. 33–50, Nov. 2018.
- [4] H. Tullberg, P. Popovski, Z. Li, M. A. Uusitalo, A. Hoglund, O. Bulakci, M. Fallgren, and J. F. Monserrat, "The METIS 5G system concept: Meeting the 5G requirements," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 132–139, Dec. 2016.
- [5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.
- [6] Y. C. *et al.*, "Toward the standardization of non-orthogonal multiple access for next generation wireless networks," *IEEE Commu. Mag.*, vol. 56, no. 3, pp. 19–27, Mar. 2018.
- [7] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

- [8] Z. Ding and V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.
- [9] J. Park and B. Clerckx, "Multi-user linear precoding for multi-polarized massive MIMO system under imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 05, pp. 2532–2547, May 2015.
- [10] T. Kim, B. Clerckx, D. J. Love, and S. J. Kim, "Limited feedback beamforming systems for dual-polarized MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, Nov. 2010.
- [11] C. Oestges, V. Erceg, and A. J. Paulraj, "Propagation modeling of MIMO multipolarized fixed wireless channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 644–654, May 2004.
- [12] X. Cheng, Y. He, and M. Guizani, "3-D geometrical model for multi-polarized MIMO systems," *IEEE Access*, vol. 5, pp. 11 974–11 984, 2017.
- [13] X. Cheng and Y. He, "Geometrical model for point-to-point multi-polarized massive MIMO systems," in *14th Int. Wireless Commun. Mob. Comput. Conf. (IWCMC)*, Jun. 2018, pp. 435–439.
- [14] F. B. *et al.*, "Multiple-antenna techniques in LTE-advanced," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 114–121, Mar. 2012.
- [15] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, Mar. 2013.
- [16] F. Zheng, Y. Chen, B. Pang, C. Liu, S. Wang, D. Fan, and J. Zhang, "An efficient CSI feedback scheme for dual-polarized massive MIMO," *IEEE Access*, vol. 6, pp. 23 420–23 430, 2018.
- [17] V. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and O. Shin, "Precoder design for signal superposition in MIMO-NOMA multicell networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2681–2695, Dec. 2017.
- [18] D. Zhang, Y. Liu, Z. Ding, Z. Zhou, A. Nallanathan, and T. Sato, "Performance analysis of non-regenerative massive-MIMO-NOMA relay systems for 5G," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4777–4790, Nov. 2017.
- [19] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Cooperative non-orthogonal multiple access in multiple-input-multiple-output channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2068–2079, Mar. 2018.
- [20] X. Chen, R. Jia, and D. W. K. Ng, "The application of relay to massive non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5168–5180, Nov. 2018.
- [21] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, "On optimal power allocation for downlink non-orthogonal multiple access systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2744–2757, Dec. 2017.
- [22] J. Wang, Q. Peng, Y. Huang, H. Wang, and X. You, "Convexity of weighted sum rate maximization in NOMA systems," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1323–1327, Sep. 2017.
- [23] L. Wu, J. Chen, H. Yang, and D. Lu, "Codebook design for LTE-A downlink system," in *IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2011, pp. 1–5.
- [24] T. Kim, B. Clerckx, D. J. Love, and S. J. Kim, "Limited feedback beamforming codebook design for dual-polarized MIMO channels," in *IEEE Global Commun. Conf. (GLOBECOM)*, Nov. 2008, pp. 1–5.
- [25] X. Su, P. Nkurunziza, J. Gu, A. Castiglione, and C. Choi, "Inter-beam interference cancellation and physical layer security constraints by 3D polarized beamforming in power domain NOMA systems," *IEEE Trans. Sust. Comp.*, pp. 1–1, 2018.
- [26] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [27] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing - The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

- [28] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Wiley Series in Probability and Statistics, Aug. 2003.
- [29] P. R. Krishnaiah and M. M. Rao, "Remarks on a multivariate Gamma distribution," *The American Mathematical Monthly*, vol. 68, no. 9, pp. 342–346, Apr. 1961.
- [30] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, 10th ed., ser. 55. USA: National Bureau of Standards, 1972.
- [31] H. Bateman and A. Erdlyi, *Higher transcendental functions*. California Institute of Technology Bateman Manuscript Project, New York: McGraw-Hill, 1953-1955, 1955, vol. 2.
- [32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., D. Zwillinger and V. H. Moll, Eds. Academic Press, 2007.