Resource Allocation in 5G Systems with Massive MIMO and NOMA Schemes

Final year project presented to the Electrical Engineering Department at Universidade Estadual de Londrina (UEL) as a requirement for the conclusion of the Bachelor of Electrical Engineering (BE) degree.

Londrina, PR
2018
Karina Bernardino Rosa

Resource Allocation in 5G Systems with Massive MIMO and NOMA Schemes

Final year project presented to the Electrical Engineering Department at Universidade Estadual de Londrina (UEL) as a requirement for the conclusion of the Bachelor of Electrical Engineering (BE) degree.

Area: Telecommunication Systems

Supervisor:
Prof. Dr. Taufik Abrão

Londrina, PR
2018
Karina Bernardin Rosa

Resource Allocation in 5G Systems with Massive MIMO and NOMA Schemes

Final year project presented to the Electrical Engineering Department at Universidade Estadual de Londrina (UEL) as a requirement for the conclusion of the Bachelor of Electrical Engineering (BE) degree.

Area: Telecommunication Systems

Examination Board

Prof. Dr. Taufik Abrão
Department of Electrical Engineering
Universidade Estadual de Londrina
Supervisor

Prof. Dr. Jose Carlos Marinello Filho
Department of Electrical Engineering
Universidade Estadual de Londrina

Prof. Msc. Jaime Laelson Jacob
Department of Electrical Engineering
Universidade Estadual de Londrina

November 17, 2018
Acknowledgments

First, I would like to thank my family for their support and encouragement during my studies, and especially my parents, for helping me through the difficult times before and during the engineering course. I also express my gratitude for my mentor during this work, Taufik Abrão. I thank him for his guidance, comments and advice, and for encouraging me. In addition, I am thankful to all the professors and friends who have helped me during the last five years.
Abstract

In this work, the downlink (DL) of a single cell MIMO system was utilized in order to analyze two distinct scenarios. First, the concept of non-orthogonal multiple access (NOMA) was applied in a cluster-based MIMO-NOMA configuration, which was evaluated in terms of spectral efficiency (SE) and resource efficiency (RE) through user grouping and a two-step power allocation strategy. The performance of the system for different beamforming directivities and cluster-loading conditions was assessed by varying the number of both clusters and users per cluster. The results demonstrated that the best performance occurs for two users per cluster, where these users are multiplexed according to the principles of NOMA. The numerical simulations also verified that the SE and EE are not monotonically decreasing due to the increase in the number of users per cluster and the simultaneous reduction in the number of transmit antennas. In the second part of this monograph, the sum-rate capacity achieved by three low-complexity linear precoders was analyzed for a Massive MIMO condition. The performance-complexity tradeoff of the conventional zero-forcing (ZF) beamforming and the regularized channel inversion (RCI) was compared to the innovative precoding scheme based on the iterative randomized Kaczmarz algorithm (rKA). A large-system limit analysis was also developed for both ZF and RCI precoding schemes. The rKA’s formulation and convergence in the DL Massive MIMO were explored, while a comprehensive complexity analysis comparing the three precoding schemes indicated the rKA’s performance-complexity tradeoff superiority.

Keywords: Channel Inversion, Kaczmarz Algorithm, Massive MIMO, NOMA, Precoding, Sum-Rate Capacity, Zero Forcing
List of Contents

Acronyms

Conventions and Notations

List of Symbols

1 Introduction 1

2 Results 3

3 Conclusions 5

Appendix A – Resource Efficiency in Cluster-Based Downlink MIMO-NOMA Systems 7

Appendix B – Sum-Rate Capacity in DL Massive MIMO with Partial CSI and Low-Complexity Linear Precoders 14

Bibliography 30
<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>5G</td>
<td>Fifth Generation</td>
</tr>
<tr>
<td>AWG</td>
<td>Additive White Gaussian</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>DL</td>
<td>Downlink</td>
</tr>
<tr>
<td>DPC</td>
<td>Dirty Paper Coding</td>
</tr>
<tr>
<td>EE</td>
<td>Energy Efficiency</td>
</tr>
<tr>
<td>EPA</td>
<td>Equal Power Allocation</td>
</tr>
<tr>
<td>IUI</td>
<td>Inter-User Interference</td>
</tr>
<tr>
<td>KA</td>
<td>Kaczmarz Algorithm</td>
</tr>
<tr>
<td>LB</td>
<td>Lower Bound</td>
</tr>
<tr>
<td>MF</td>
<td>Matched-Filtering</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple Input Multiple Output</td>
</tr>
<tr>
<td>MRC</td>
<td>Maximum-Ratio Combining</td>
</tr>
<tr>
<td>MT</td>
<td>Mobile Terminal</td>
</tr>
<tr>
<td>MU</td>
<td>Multi-User</td>
</tr>
<tr>
<td>NOMA</td>
<td>Non-Orthogonal Multiple Access</td>
</tr>
<tr>
<td>OD</td>
<td>Overdetermined</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>OMA</td>
<td>Orthogonal Multiple Access</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>RCI</td>
<td>Regularized Channel Inversion</td>
</tr>
<tr>
<td>RE</td>
<td>Resource Efficiency</td>
</tr>
<tr>
<td>rKA</td>
<td>Randomized Kaczmarz Algorithm</td>
</tr>
<tr>
<td>SC</td>
<td>Superposition Coding</td>
</tr>
<tr>
<td>SE</td>
<td>Spectral Efficiency</td>
</tr>
<tr>
<td>SIC</td>
<td>Successive Interference Cancelation</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-Interference-plus-Noise Ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single Input Single Output</td>
</tr>
<tr>
<td>SLE</td>
<td>Set of Linear Equations</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>UB</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>UD</td>
<td>Underdetermined</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
<tr>
<td>ULA</td>
<td>Uniform Linear Array</td>
</tr>
<tr>
<td>UPA</td>
<td>Uniform Planar Array</td>
</tr>
<tr>
<td>ZF</td>
<td>Zero-Forcing</td>
</tr>
</tbody>
</table>
Conventions and Notations

The following mathematical notations were adopted in this work:

**a** Boldface lower case letters represent vectors;

**A** Boldface upper case letters denote matrices;

$(·)^{-1}$ Inversion operator;

$(·)^H$ Hermitian operator (transposition and conjugation);

$(·)^T$ Transposition operator;

$(·)^*$ Conjugation operator;

$(·)^*$ Optimal solution;

$\det(·)$ Determinant of a Square Matrix;

$\|·\|_n$ Norm of order $n$;

$\|·\|_F$ Frobenius norm;

$\text{tr}(·)$ Trace operation;

$\text{diag}(·)$ Diagonalization Operation;

$I_m$ Identity matrix of order $m$;

$0_{m \times n}$ All zero matrix of size $m \times n$;

$\mathcal{CN}\{\mu, \sigma^2\}$ Gaussian Random Variable circularly-symmetric with mean $\mu$ and variance $\sigma^2$;

$\mathcal{O}(·)$ Complexity order of an operation or algorithm;

$\mathbb{E}[·]$ Statistical Expectation;

$\text{Var}[·]$ Variance;

$\mathbb{C}$ Complex numbers set;
\( \Re \{ \cdot \} \) Real part of a complex number;

\( \Im \{ \cdot \} \) Imaginary part of a complex number;

\( \mathcal{G}^t(\cdot) \) Linear operator that depends on the internal randomization of KA until iteration \( t \)

\( \in \) Belongs to the set;

\( \langle \mathbf{X}, \mathbf{Y} \rangle \) Inner product between two matrices or vectors \( \mathbf{X} \) and \( \mathbf{Y} \)
## List of Symbols

### Appendix A

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>Number of antennas at the BS</td>
</tr>
<tr>
<td>$K$</td>
<td>Number of single-antenna users (receivers)</td>
</tr>
<tr>
<td>$L$</td>
<td>Number of clusters</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Number of receivers in the $m$-th cluster</td>
</tr>
<tr>
<td>$K_c$</td>
<td>Fixed number of receivers in each cluster</td>
</tr>
<tr>
<td>$r$</td>
<td>Cell radius</td>
</tr>
<tr>
<td>$r_1$</td>
<td>Smaller cell disk radius</td>
</tr>
<tr>
<td>$s$</td>
<td>Symbol vector</td>
</tr>
<tr>
<td>$x$</td>
<td>Transmit signal vector</td>
</tr>
<tr>
<td>$A$</td>
<td>Path-loss coefficient matrix</td>
</tr>
<tr>
<td>$b$</td>
<td>Path-loss exponent</td>
</tr>
<tr>
<td>$H$</td>
<td>Channel matrix</td>
</tr>
<tr>
<td>$\tilde{H}$</td>
<td>Equivalent channel matrix</td>
</tr>
<tr>
<td>$H_m$</td>
<td>Channel matrix of the $m$-th cluster</td>
</tr>
<tr>
<td>$\tilde{H}_m$</td>
<td>Equivalent channel vector of the $m$-th cluster</td>
</tr>
<tr>
<td>$z$</td>
<td>AWG noise vector</td>
</tr>
<tr>
<td>$y$</td>
<td>Received signal vector</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Noise variance</td>
</tr>
<tr>
<td>$G$</td>
<td>Linear precoding matrix</td>
</tr>
<tr>
<td>$P$</td>
<td>Diagonal power matrix</td>
</tr>
<tr>
<td>$P_T$</td>
<td>Total available transmit power at the BS</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>SNR at the transmitter side</td>
</tr>
<tr>
<td>$\text{SINR}_{m,k}$</td>
<td>SINR of the $k$-th user at the $m$-th cluster</td>
</tr>
<tr>
<td>$g_{m,k}$</td>
<td>Channel gain of the $k$-th user at the $m$-th cluster</td>
</tr>
<tr>
<td>$R_{m,k}$</td>
<td>Capacity of the $k$-th user at the $m$-th cluster</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Power normalization constant</td>
</tr>
<tr>
<td>$\xi$</td>
<td>Regularization parameter</td>
</tr>
<tr>
<td>$\text{SE}$</td>
<td>Spectral Efficiency</td>
</tr>
<tr>
<td>$\text{EE}$</td>
<td>Energy Efficiency</td>
</tr>
<tr>
<td>$\text{RE}$</td>
<td>Resource Efficiency</td>
</tr>
<tr>
<td>$P_c$</td>
<td>Constant circuit power consumption per antenna</td>
</tr>
<tr>
<td>$P_0$</td>
<td>Basic power consumed at the BS</td>
</tr>
</tbody>
</table>
Constant accounting for the inefficiency of the power amplifier at the BS

Appendix B

\( M \)  
Number of antennas at the BS

\( K \)  
Number of single-antenna users (receivers)

\( \beta \)  
Cell-loading factor

\( s \)  
Symbol vector

\( x \)  
Transmit signal vector

\( A \)  
Path-loss coefficient matrix

\( r_k \)  
Distance between the BS and the \( k \)-th user

\( b \)  
Path-loss exponent

\( H \)  
Channel matrix

\( Q \)  
Estimated channel matrix

\( \tau \)  
Channel estimation quality parameter

\( n \)  
AWG noise vector

\( y \)  
Received signal vector

\( \sigma_n^2 \)  
Noise variance

\( G \)  
Linear precoding matrix

\( P \)  
Diagonal power matrix

\( P_t \)  
Total available transmit power at the BS

\( \gamma \)  
SNR at the transmitter side

\( \Phi \)  
Transmit correlation matrix

\( c \)  
Correlation parameter

\( R_k \)  
Capacity of the \( k \)-th user

\( R_\Sigma \)  
\textit{Ergodic} sum-rate capacity

\( G \)  
Precoding matrix

\( \alpha \)  
Power normalization constant

\( \text{SINR}_k \)  
SINR of the \( k \)-th user

\( \xi \)  
Regularization parameter

\( \rho \)  
Normalized regularization parameter in the Large System Limit

\( S_k \)  
Signal of the \( k \)-th user in the Large System Limit

\( I_k \)  
Interference of the \( k \)-th user in the Large System Limit

\( \text{SINR}^\infty \)  
Limiting SINR of the \( k \)-th user

\( R_k^\infty \)  
Limiting capacity of the \( k \)-th user

\( R_\Sigma^\infty \)  
Limiting sum rate capacity

\( p \)  
Row probability distribution vector
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>Number of rKA iterations</td>
</tr>
<tr>
<td>$\mathbf{T}_{\text{all}}$</td>
<td>Interference matrix</td>
</tr>
<tr>
<td>$\underline{R}_k$</td>
<td>Lower bound on the achievable ergodic rate of the $k$-th user</td>
</tr>
<tr>
<td>$\overline{R}_k$</td>
<td>Upper bound on the achievable ergodic rate of the $k$-th user</td>
</tr>
<tr>
<td>$T_{\text{full convg}}$</td>
<td>Number of iterations for the full convergence of rKA</td>
</tr>
</tbody>
</table>
1 Introduction

Fifth generation (5G) communication systems are expected to bring several advances and benefits, such as higher data transmission rates, high reliability and latency reduction. The systems will also be more connected, using a massive number of transmitters which seek to deliver a homogeneous connection to the numerous users, regardless of their location or network traffic conditions.

Therefore, the 5G wireless communications need to meet challenging requirements such as high spectral and energy efficiency, as well as massive connectivity. To meet these requirements, several technologies have been proposed and exploited, such as massive multiple-input-multiple-output (MIMO). In the majority of the MIMO implementations, the base station (BS) employs a small number of antennas, and the corresponding improvement in spectral efficiency, while important, is still relatively modest. The use of MIMO systems with a huge number of antennas, which is called massive MIMO, is a solution to solve these problems related to the need for high data rates, simultaneously serving many terminals in the same time-frequency resource and attaining the benefits of conventional MIMO, but on a larger scale (LU et al., 2014; G. et al., 2014).

In general, massive MIMO presents significant advantages in terms of energy efficiency, spectral efficiency, robustness and reliability. Nonetheless, problems such as pilot contamination and high computational complexity need to be addressed in order to fully realize the potential of the technology. In particular, the precoding and detection processes impose a significantly high computational complexity on massive MIMO implementations. In the last years, a significant research effort has been dedicated to developing low-complexity algorithms for the design of precoding/detection matrices. Most of the works, however, require statistical information of the users channels to reduce the complexity, what is also a computationally demanding problem and may not be practical, especially in massive MIMO regime. Hence, new methods, topologies, and/or algorithmic solutions for massive MIMO are expected to be efficient and have the lowest computational complexity possible (G. et al., 2014; BOROUJERDI; HAGHIGHATSHOAR; CAIRE,
In 5G communication networks, new multiple access technologies are also important for the massive connectivity of a large number of mobile users and devices connected to the Internet of Things (IoT) with different quality of service (QoS) requirements. A technology proposed for meeting these requirements is the use of non-orthogonal multiple access (NOMA) schemes. Over the history of wireless communications, the applied multiple access scheme has been a relevant distinction between different systems. In many of the conventional multiple access schemes, such as time division multiple access (TDMA) and orthogonal frequency division multiple access (OFDMA), resource allocation is done orthogonally over time or frequency so as to avoid interference between users. However, this type of scheme presents problems such as the fact that resources occupied by users with poor channel conditions, which are therefore not fully exploited, cannot be shared with other users (WONG et al., 2017).

In contrast to these conventional technologies of orthogonal multiple access (OMA), NOMA schemes can accommodate many more users through non-orthogonal resource allocation (DAI LINGLONG; WANG et al., 2015). NOMA resource allocation schemes allow users to share the same spectrum and time resources by allocating them to different codes or different power levels. NOMA techniques which exploit different power levels for user multiplexing, i.e., Power Domain NOMA, are able to apply the available resources in a more efficient way, capitalizing the specific channel conditions of each user and serving multiple users with different QoS using the same spectrum, time and code resources (LIU et al., 2017; DING et al., 2017).
2 Results

The numerical simulation results, the associated analyses, as well as the analytical analyses related to both parts of this work are presented herein in the form of articles written throughout the development of the activities. Such papers have been submitted to conferences and for a journal. In the sequel, these articles are briefly introduced and appended as part of this document.

[A] Resource Efficiency in Cluster-Based Downlink MIMO-NOMA Systems
Authors: Karina Bernardin Rosa, Taufik Abrão
Status: Under review (Nov. 29th, 2018)

[B] Sum-Rate Capacity in DL Massive MIMO with Partial CSI and Low-Complexity Linear Precoders
Authors: Karina Bernardin Rosa, João Lucas Negrão, Jose Carlos Marinello, Taufik Abrão
Paper submitted to the journal Transactions on Emerging Telecommunications Technologies (ETT).
Status: Under review (Nov. 29th, 2018)

From [B], two short-papers were also generated and submitted to conferences, as follows:

[C] Correlation and Partial CSI on the DL Massive MIMO Low-Complexity Kaczmarz Precoding
Authors: Karina Bernardin Rosa, Jose Carlos Marinello, Taufik Abrão
Status: Under review (Nov. 29th, 2018)
[D] *Low Complexity Kaczmarz Precoding for Massive MIMO Under Correlation and Partial CSI*

Authors: Karina Bernardin Rosa, Jose Carlos Marinello, Guixian Xu, Taufik Abrão

Paper submitted to the conference IEEE ICC 2019 (IEEE International Conference on Communications).

Status: Under review (Nov. 29th, 2018)

In [A], a cluster-based MIMO-NOMA system configuration is analyzed in terms of spectral and resource efficiency through a two-step power allocation strategy. The performance of the system for different beamforming directivities and cluster-loading conditions was assessed through variation in the number of both clusters and users per cluster. The work in [B] analyzes and compares the sum-rate capacity achieved by three low-complexity linear precoders in a single cell Massive MIMO broadcast channel; these linear precoders are the conventional zero-forcing beamforming, the regularized channel inversion and a precoding scheme based on the iterative randomized Kaczmarz algorithm. A large-system limit analysis is presented for both ZF and RCI. The rKA’s formulation and its convergence for the DL is explored, and a complexity analysis of the three precoding schemes is presented and indicates the rKA’s performance-complexity trade-off superiority compared to the other conventional techniques when operating in Massive MIMO systems. As mentioned above, [C] and [D] were derived from [B], presenting a compact version of this article’s results and some new discussions regarding the sum-rate capacity, convergence and complexity of the rKA precoder.
3 Conclusions

Through this final year project, several concepts and issues associated to DL massive MIMO scenarios were analyzed; specifically, we have focused on exploratory numerical analyses of the power domain applied to MIMO user multiplexing, the performance analysis of massive MIMO precoding schemes and their associated computational complexity.

Initially, the spectral and resource efficiencies of a cluster-based MIMO-NOMA system configuration were evaluated in order to determine the influence of the number of antennas, as well as the number of users multiplexed in power, in the system performance. For assessment purposes, we have considered a simple user grouping and a two-step power allocation strategy; the best performance of the system was achieved for a condition with the highest number of antennas at the BS and only two users sharing the same power resources. The following best performances, however, were not attained respectively for the following conditions of least amount of users per cluster. Instead, it was verified that in some configurations a higher number of users in a cluster was shown to have a superior performance. Besides, the analysis of the system demonstrated, as expected, that an excessive number of users multiplexed in power can result in performance degradation.

The second part of this work focused on the sum-rate capacity analysis achieved by three low-complexity linear precoders in a massive MIMO scenario equipped with uniform linear array (ULA) antennas, and considering the impact of channel correlation, as well as the imperfect channel state information on the system performance-complexity tradeoff. The conventional ZF and RCI precoders were compared with the iterative randomized Kaczmarz algorithm (rKA) precoding scheme in terms of sum-rate capacity and complexity. Numerical results demonstrated that, despite the fact the number of iterations needed for the convergence of rKA is reduced for poor channel estimation quality, it increases with the SNR, spatial channel correlation and cell load conditions. Even though, its performance-complexity tradeoff is still more favorable than the conventional
RCI and ZF precoding when operating in realistic massive MIMO scenarios and conditions, what is significantly due to the low-complexity of the rKA precoding algorithm, which does not depend on the channel inversion. The rKA precoder also demonstrated to be equally efficient if compared to RCI under a simple equal power allocation policy and an adequate number of iterations. Moreover, in this work, it was shown that imperfect channel estimation provoked a similar performance degradation for the three precoding techniques.
Appendix A – Resource Efficiency in Cluster-Based Downlink MIMO-NOMA Systems
Resource Efficiency in Cluster-Based Downlink MIMO-NOMA Systems

Karina Bernardin Rosa, Taufik Abrão
Department of Electrical Engineering, State University of Londrina
Po.Box 10.011, CEP: 86057-970, Londrina, PR - Brazil
karinabernardin@gmail.com, taufik@uel.br

Abstract—This work discusses a cluster-based non-orthogonal multiple access (NOMA) structure aiming at improving the resource efficiency, namely the energy and spectral efficiencies (EE and SE), in a cell through power allocation. In the considered scenario, the users are grouped into several clusters; hence, each beamformer serves the users located at a specific cluster. The RE in a cluster-based downlink (DL) MIMO NOMA is attained through the application of a two-step simple power allocation strategy. The cluster head in each cluster is able to almost completely cancel the intra-cluster interference. Our results demonstrated that the best performance occurs for two users per cluster.

I. INTRODUCTION

In contrast to the conventional orthogonal multiple access (OMA), the NOMA can fairly accommodate more than one user via non-orthogonal resource allocation, via power-domain multiplexing, with exploitation of the channel differences among users, features that were not exploited in current and past cellular systems. The large power difference to deal with cell-center and cell-edge users facilitates the successful decoding of the signals designated for each user, enabling the use of relatively low-complexity receivers at the receiver side [1][2].

The NOMA schemes are expected to increase the system throughput. Beyond that, user fairness and spectral efficiency can be significantly improved under NOMA schemes, since the users, served by different power levels, are able to transmit at the same frequency and at the same time slot [1]. A possibility for improving spectral efficiency with NOMA is exploiting also the spatial degrees of freedom enabled by MIMO technologies, which is crucial for complying the performance requirements of 5G networks [3].

Multiple-antenna-aided NOMA design is capable of using directional beamforming or spatial multiplexing in order to provide array gains or increase the system’s throughput, respectively. Also, adopting appropriate transmit precoding matrix design, enables to create user-specific channels. Exploring these configurations can lead to a generalized NOMA design for satisfying the heterogeneous QoS user requirements [4].

The fast growth of data traffic in the recent systems has been accompanied with an increase in the energy consumption. This situation, combined to limitations in the battery capacity occasioned by reduction in the terminal size limits, makes the energy efficiency (EE) one dominant target in the design of future wireless communication systems. As a consequence, in the recent years numerous works focusing on analyzing and enhancing not only the spectral efficiency (SE), but also the EE of wireless communication systems. As a consequence, in the recent years numerous works focusing on analyzing and enhancing not only the spectral efficiency (SE), but also the EE of wireless communication systems.

Different from the most of works, which only consider either SE or EE, authors in [5] present a more general and flexible measure, called the Resource Efficiency (RE). The RE concept has also been explored in [6] for massive MIMO applications; RE is defined as a weighted combination of EE and SE. This new performance measure has more flexibility in striking the balance between SE and EE, despite difficulties in solving the resulting optimization problem. Hence, authors first investigate an uplink-downlink duality for the RE maximization purpose. The conventional UL-DL duality only applies in the SE or the capacity region, and has yet not been explicitly established for the EE or the RE. In [6], authors prove with rigorous derivation that the duality has a more general form which can be directly used to tackle either the EE or RE beamforming design problem.

According to [4], current downlink MIMO-NOMA applications can be broadly classified into two categories: the beamformer-based structure, in which one beamformer is used for each user, and the cluster-based structure, on which one beamformer serves multiple users. [4]. In this work we focus on the cluster-based structure, which is based on separating the users and grouping them into several clusters, while each beamformer serves all the users of a specific cluster. Then, by applying appropriate detector design, the inter-cluster interference can be reduced substantially or even totally suppressed, depending on which type of beamformer is used.

Decomposing techniques capable to transform MIMO-NOMA channels into multiple SISO-NOMA channels are discussed in [7], [8]. The authors consider a BS equipped with M antennas while the users equipped with N antennas each are randomly partitioned into M clusters. The detection technique used in these designs is then based on the zero forcing (ZF) detection, and TPC designs can be applied in order to completely cancel the intercluster interference. Thus, this approach leads to low-complexity SISO-NOMA model where the conventional NOMA technique can be applied. Despite of the clear advantages in using this proposed design, the zero-forcing detection adopted with such techniques requires a number of antennas at each receiver (N) be greater than or equal the number of transmit antennas (N ≥ M) [7] or at least to be greater than or equal half the number of transmit antennas (N ≥ M/2) [8].

Another cluster-based MIMO-NOMA design is proposed by [9]. In contrast to the ones mentioned before, this design allows the existence of intercluster interference, applying specific user grouping and power allocation schemes in order to reduce this kind of interference. Besides, it was assumed a configuration with N transmit antennas at the BS and N clusters, with the assumption that each cluster has two users for the sake of simplicity. It was also assumed each user has a single antenna. Among the users in a specific cluster, the user having a larger (smaller) channel gain is defined as a strong (weak) user. As a simple and practical alternative, zero-forcing beamforming (ZF-BF) is used under perfect channel state information at the transmitter (CSIT), and the precoding is performed by considering the channel gain of a particular user of each cluster. The user using the BF vector based on its own channel does not receive any interference from the other beams, i.e., inter-cluster interference is virtually zero. On the other hand, the other
users belong to the same cluster suffer inter-cluster interference, which negatively affects the decoding of the received signal. Since it is unhelpful when the strong users perform successive interference cancelation (SIC), the authors consider the generation of the BF vectors based on the channels of the strong users of each cluster. In order to minimize the inter-cluster interference of the weaker channel users and maximize the sum capacity while guaranteeing the capacity requirement of the weak user, the authors respectively propose a clustering algorithm and a suitable power allocation scheme.

Another cluster-based MIMO-NOMA design is proposed in [10]. This design also allows the existence of intercluster interference, applying a technique for reducing the interference and increase the strength of the desired signals. The authors of [10] also utilize a ZF-BF technique, but in this approach the precoding is performed by considering the equivalent channel gain of each cluster, instead of any particular user channel gain.

Instead of using random clustering as in [7] and [8], this user-clustering approach is based on making the channel gains of users more distinctive and sorting the users in a specific cluster according to their equivalent normalized channel gains. This strategy enables the user with the highest channel gain, namely cluster head, to have its gain very similar to the cluster equivalent channel gain, and thus to be able to almost completely cancel the inter-cluster interference and completely cancel intra-cluster interference by invoking ideal SIC. Each of the other users in the cluster, then, efficiently suppress the inter-cluster interference by estimating their own cluster’s equivalent channel gain by multiplying their received signal by a user-specific decoding scaling weight factor sent by the base station (BS) prior to the data transmission process.

The main advantage of these two inter-cluster interference-tolerant designs is that they do not impose any number of antennas constraints at the BS or at the users, which implies that the number of antennas in each user is not somehow attached to the number of antennas at the BS and, thus, these designs can be applied to scenarios where the BS is equipped with a large antenna array, such as in massive-MIMO-NOMA or NOMA millimeter-wave communication scenarios.

In this paper, we consider the multiuser power allocation for a cluster-based downlink MIMO-NOMA system and aim at achieving the balance between the SE and the EE through the use of the RE as a more general and flexible measure.

The main contributions of this work are threefold. We have analyzed a clustering-based non-orthogonal multiple access (NOMA) structure under the resource efficiency (RE) perspective, which combines the energy and spectral efficiencies (EE and SE) performance metrics; we have demonstrated that there is a preferable grouping strategy for achieving the RE optimality; finally, this work evaluates the computational complexity of the grouping-based NOMA-MIMO schemes from the perspective of the performance-complexity tradeoff.

Notation. In the following, boldface lower-case and uppercase characters denote vectors and matrices, respectively. The operators $(\cdot)^H$, tr $(\cdot)$ and $\mathbb{E} [\cdot]$ denote conjugate transpose, trace and expectation, respectively. The $M \times M$ identity matrix is denoted by $I_M$. A random vector $x \sim \mathcal{CN}(\mathbf{m}, \Theta)$ is complex Gaussian distributed with mean vector $\mathbf{m}$ and covariance matrix $\Theta$.

The remainder of this paper is divided as follows. The system model is described in Section II. NOMA precoding designs and user grouping are discussed in Section III. In Section IV, we discuss the RE concept. Power allocation schemes are discussed in Section V. In Section VI, numerical results are provided, and the work is concluded in Section VII.

II. SYSTEM MODEL

In this work we consider a downlink multi-user (MU) MIMO-NOMA system with a single cell and one BS equipped with $M$ transmit antennas used for beamforming. The total number of user equipments (UE) in a cell is $K$, where $K \geq M$, and each UE is considered to have one receive antenna. Also, the receive antennas are grouped into $L$ clusters. The users in each cluster are scheduled according to the NOMA principles.

In the assumed conditions $L = M$ and each BF vector serves an individual cluster, which means all clusters use the same spectrum resources. The $m$-th cluster consists of $K_m$ receive antennas such that $\sum_{m=1}^{M} K_m = K$.

In practice, it may not be realistic to schedule all the users in a cluster using NOMA. An alternative is to build a hybrid system, in which NOMA is combined with OMA techniques such as OFDMA. In this condition, $L > M$ and multiple clusters can utilize the same BF vector while using orthogonal spectrum resources to each other, while the users in each cluster are scheduled according to the NOMA principles.

For simplicity, in this analysis we assume the number of clusters equal to the number of transmit antennas, each UE equipped with a single antenna and a fixed number $K$ of users in each cluster, i.e., $L = M$ and $K = K \times M$.

The users are considered to be randomly distributed an a disk, and the BS is assumed to be located at its center. The radius of this disk is denoted by $r$. We then divide this disk in two parts: one smaller disk, with a radius $r_1$ ($r_1 < r$), and with the BS located at its center. The other part is a ring, constructed from the larger disk by removing the smaller one. Fig. 1 illustrates a generic system model which follows the described assumptions.

![Figure 1. Generic system model.](image)

We consider a channel model with Rayleigh fading and large scale path loss.

The data stream for the $m$-th cluster, corresponding to the $m$-th element of the transmitted data vector is defined as:

$$s_m = \sum_{k=1}^{K} \sqrt{p_{m,k}s_{m,k}},$$

(1)

where $p_{m,k}$ and $s_{m,k} \sim \mathcal{CN}(0,1)$ are the transmit power and symbol of the $k$-th user from the $m$-th cluster, and $\sum_{k=1}^{K} p_{m,k} = 1$.

The transmitted data vector $\mathbf{s} \in \mathbb{C}^{M \times 1}$ is then defined as:

$$\mathbf{s} = [s_1 \ s_2 \ \ldots \ s_m \ \ldots \ s_M]^\top.$$  

(2)

Also, the data vector $\mathbf{s}$ is multiplied by a power matrix $\mathbf{P} \in \mathbb{R}^{M \times M}$ and a precoding matrix $\mathbf{G} \in \mathbb{C}^{M \times M}$ and, then, transmitted over a radio channel $\mathbf{H} \in \mathbb{C}^{K \times M}$. Each $\mathbf{H}_m \in \mathbb{C}^{K \times M}$ represents
the radio channel of all $K$ users in the $m$-th cluster, and can be expressed as:
\[
H_m = [h_{m,1} \ h_{m,2} \ \ldots \ h_{m,K}]^T, \tag{3}
\]
with $h_{m,k} \in \mathbb{C}^{1 \times M}$ as the radio channel gain vector of the $k$-th user in the $m$-th cluster.

Thus, the radio channel matrix $H \in \mathbb{C}^{K \times M}$ is expressed as:
\[
H = [H_1 \ H_2 \ \ldots \ H_m \ \ldots \ H_M]^T. \tag{4}
\]

The power matrix corresponding to the power allocated to each cluster is defined as $P = \text{diag}(p_1, \ldots, p_M) \in \mathbb{R}^{M \times M}$, with $\sum_{m=1}^{M} p_m = P_1$. The transmitted signal $x \in \mathbb{C}^{M \times 1}$ obtained after the power and precoding multiplication can be written as:
\[
x = GP^{1/2} = \sum_{m=1}^{M} \sqrt{p_m}g_m s_m. \tag{5}
\]

Also, the precoding vectors are normalized to satisfy the average power constraint, thus:
\[
\mathbb{E} \left[ ||x||^2 \right] = \text{tr} \left( PGH G \right) \leq P_1, \tag{6}
\]
where $P_1 \geq 0$ is the total available transmit power at the BS, hence, the SNR at the receiver side is denoted as $\gamma = \frac{P_1}{\sigma^2}$.

The received signal $y \in \mathbb{C}^{K \times 1}$ can be expressed as:
\[
y = [y_1 \ y_2 \ \ldots \ y_m \ \ldots \ y_M]^T, \tag{7}
\]
where each $y_m \in \mathbb{C}^{K \times 1}$ is constituted by the signals received for each of the $K$ users in the $m$-th cluster, such as:
\[
y_m = [y_{m,1} \ y_{m,2} \ \ldots \ y_{m,k} \ \ldots \ y_{m,K}]^T, \tag{8}
\]
where $y_{m,k} \in \mathbb{C}$ corresponds to the signal received by the $k$-th user of the $m$-th cluster.

The array $y \in \mathbb{C}^{K \times 1}$ can be written in terms of its components as:
\[
y = HGP^{1/2}s + z, \tag{9}
\]
where $z \in \mathbb{C}^{K \times 1}$ represents the complex gaussian noise vector with variance $\sigma^2$, whose elements are represented as $z_{m,k} \in \mathbb{C} \sim \mathcal{CN}(0, \sigma^2)$. The received signal for the $k$-th user in the $m$-th cluster can then be expressed as:
\[
y_{m,k} = h_{m,k}GP^{1/2}s + z_{m,k}. \tag{10}
\]

III. PRECODING DESIGN AND GROUPING STRATEGY

In the MIMO-NOMA model under consideration $K > M$, and thus we utilize a precoding technique suggested by [10] based on [11], where the actual channel matrix $H_m \in \mathbb{C}^{K \times M}$ corresponding to the $K$ users of the $m$-th cluster is transformed into an equivalent channel vector $h_m \in \mathbb{C}^{1 \times M}$. This manipulation then leads to a total channel matrix $H \in \mathbb{C}^{M \times M}$ as an equivalent to $H_m$, which provides compatible dimension for precoding application. The radio channel matrix corresponding to the $m$-th cluster can be expressed as:
\[
H_m = [h_{m,1} \ h_{m,2} \ \ldots \ h_{m,k} \ \ldots \ h_{m,K}]^T. \tag{11}
\]

Taking the singular value decomposition (SVD) of $H_m$, we obtain:
\[
H_m^{[K \times M]} = U_m^{[K \times K]}\Sigma_m^{[K \times M]} V_m^{H^{[M \times M]}}, \tag{12}
\]
In the considered system, each beamforming vector is utilized by one cluster. According to this configuration, the equivalent radio channel matrix $\hat{H}_m \in \mathbb{C}^{1 \times M}$ representing the equivalent channel of the $m$-th cluster can be obtained as:
\[
\hat{H}_m^{[1 \times M]} = U_m^{H^{[1 \times K]}}U_m^{[K \times K]}\Sigma_m^{[K \times M]} V_m^{H^{[M \times M]}}, \tag{13}
\]
where $U_m^{[1 \times K]}$ is the Hermitian transpose of the first column of $U_m$ in (12).

An alternative for calculating the equivalent channel matrix is to assume the channel of the user closer to the BS in each cluster as the cluster equivalent channel vector, such as adopted in [9]. According to both approaches, the equivalent radio channel matrix $\hat{H} \in \mathbb{C}^{M \times M}$ formed by the equivalent channels of all the $M$ clusters can be expressed as:
\[
\hat{H} = [\hat{h}_1 \ \hat{h}_2 \ \ldots \ \hat{h}_m \ \ldots \ \hat{h}_M]^T. \tag{14}
\]

We considered as our precoding technique the Regularized Zero-Forcing (RZF) precoding, that can be faced as a generalization of the Zero-Forcing (ZF) precoding, in which a regularization parameter is added to the pseudo-inverse matrix. Considering a system equipped with the RZF precoder and the equivalent channel matrix $\hat{H}$, the precoding matrix solution is given by [12], [13], [14]:
\[
G_{ZF} = \alpha \hat{H}^H (\hat{H}^H \hat{H} + \xi I_M)^{-1}, \tag{15}
\]
or equivalently as
\[
G_{ZF} = \alpha (\hat{H}^H \hat{H} + \xi I_M)^{-1} \hat{H}^H, \tag{16}
\]
where the normalizing constant $\alpha$ is chosen to satisfy the power constraint (6), and $\xi > 0$ is the regularization parameter.

As stated in [15], by assuming independence between the data symbols, the normalization constant for the RCI precoding is expressed as
\[
\alpha = \sqrt{\text{tr} \left( P \hat{H}^H (\hat{H}^H \hat{H} + \xi I_M)^{-2} \hat{H} \right)}, \tag{17}
\]
where now the normalization factor $\alpha$ depends on the channel realization $\hat{H}$, as well as the regularization factor $\xi$.

Using the RZF precoder (16), the received vector $y$ and signal $y_{m,k}$ for each user can be respectively expressed as
\[
y = \alpha \hat{H}^H \hat{H} + \xi I_M)^{-1} \hat{H}^H P^{1/2}s + n \tag{18}
\]
Since SIC is performed within each MIMO-NOMA cluster, the dynamic power allocation in each of them is performed in a way a user can decode and then suppress the intra-cluster interference from users with channel gains lower then its own. Thus, the signal received for the $k$-th user of the $m$-th cluster is denoted by:
\[
y_{m,k} = \alpha \sqrt{p_m} h_{m,k} (\hat{H}^H \hat{H} + \xi I_M)^{-1} h_{m,k}^H s_{m,k} + \alpha \sqrt{p_m} h_{m,k} (\hat{H}^H \hat{H} + \xi I_M)^{-1} h_{m,k}^H \sum_{j=1, j \neq k}^{M} \sqrt{p_j} s_{m,j} + \text{intra-cluster interference}
\]
\[
+ \sum_{i=1, i \neq m}^{M} \alpha \sqrt{p_i} h_{i,k} (\hat{H}^H \hat{H} + \xi I_M)^{-1} h_{i,k}^H s_i + z_{m,k} \tag{19}
\]
where the first term in the right side of (19) is the desired signal for the $k$-th user from the $m$-th cluster, while the other terms are the interference introduced by the other users in the same cluster and in other clusters plus the received thermal noise. In order to avoid
excessive complexity, we advocate the use of single-user detection at the DL receiver side. Hence, the SINR for each user is expressed as in Eq. 21 [12][16].

The achievable throughput for the \( k \)-th user of the \( m \)-th cluster can then be expressed as:

\[
R_{m,k} = \log_2 \left( 1 + \text{SINR}_{m,k} \right),
\]

where \( \eta \) is the weighting factor in \( \frac{\text{SE}}{\text{RE}} \) controlling the weights of EE and the SE on the design. Hence, when \( \eta = 0 \) the expression (23) reduces to the EE, and it tends to the SE for \( \eta \gg 1 \). The value choice for this weighting factor is then up to the system designer [6].

A. User Grouping Strategy

In a conventional NOMA system, it is preferable to pair users whose channel conditions are significantly different, what improves the sum-rate of both users and also each individual user’s rate [1]. Based on the work presented in [10], in this work we utilize a sub-optimal user clustering scheme for downlink MIMO NOMA systems, which exploits the channel gain differences among the users in a cell in order to allow an optimal power allocation scheme targeting an enhancement in the sum-spectral efficiency in the cell.

According to the user-clustering scheme presented in this work, the cluster head in each downlink MIMO NOMA cluster can almost completely cancel the intra-cluster interference, and, thus, achieves maximum throughput gain in comparison to the other users in the cluster. Based on this consideration, one strategy to maximize the overall system capacity in each cluster is to select the high channel gain users in the cell as the cluster-heads of different MIMO NOMA clusters. This consideration can also be used to determine the number of clusters in the cell, based on the number of high channel gain users available.

IV. RESOURCE EFFICIENCY

Instead of focusing on the SE or the EE separately as in the traditional design, it is much more effective balancing the attainable system SE and EE by adopting the resource efficiency (RE) metric, such as presented in [5] and adopted in [6].

Conventional system designs usually focus on the SE defined for a single cell system with \( K \) single-antenna users and \( M \) BS antennas as \( \text{SE} = \sum_{i=1}^{K} P_{i} \), with \( P_{i} = \log_2(1 + \text{SINR}_{i}) \). On the other hand, the energy efficiency (EE) has become an important figure of merit for the future wireless communications; the EE is expressed as a weighted sum of the resource efficiency in the cell.

\[
\text{EE} = \frac{\text{SE}}{\sum_{i=1}^{K} P_{i} + MP_{e} + P_{0}}, \tag{22}
\]

where \( P_{e} \) is the constant circuit power consumption per antenna including power dissipations in the transmit filter, the mixer, the frequency synthesizer, and the digital-to-analog converter, \( P_{0} \) is the basic power consumed at the BS which is independent of the number of transmit antennas and \( \varepsilon > 1 \) is a constant accounting for the inefficiency of the power amplifier at the BS [6].

The resource efficiency is expressed as a weighted sum of the EE and the SE, and can be formulated as:

\[
\text{RE} = \text{EE} + \eta \frac{\text{SE}}{\sum_{i=1}^{K} P_{i} + MP_{e} + P_{0}}, \tag{23}
\]

A. CSI Based Power Allocation

In this MIMO-NOMA system, the power allocation is performed in a two-step method, such as proposed in [10]. First, the total BS transmit power is divided into the number of transmit beams, and the transmit power for a beam is proportional to the number of users served by that beam/cluster, i.e. \( P_{m} \propto K_{m} \) where \( K_{m} \) is the number of users inside the cluster \( m \).

Finally, the RE for MIMO-NOMA is then defined combining (24) and (25) into (23) as:

\[
\text{RE}_{\text{MIMO-NOMA}} = \frac{\text{EE}_{\text{SE}}}{\sum_{m=1}^{M} P_{m} + MP_{e} + P_{0}}, \tag{26}
\]

V. POWER ALLOCATION

In this MIMO-NOMA system, the power allocation is performed in allocating power for the users in each cluster. The power allocation for the \( K \) users at the \( m \)-th cluster is scheduled according to the NOMA principle, and thus the intra-cluster dynamic power allocation is essential. In the sequel we examine the power allocation based on the inverse of channel state.

A. CSI Based Power Allocation

In this strategy, the power coefficients are determined based on the channel state information experienced by each user of the cluster. The relation between the power shared per user and its channel state is assumed to be inversely proportional, given by [17]

\[
p_{m,k} \propto \frac{1}{g_{m,k}}, \tag{27}
\]
which means the BS assigns highest power fractions to users which experience bad channel conditions, while assigning less power to the users with better channel conditions. The power fraction assigned to the $k$-th user from $m$-th cluster is given by:

$$p_{m,k} = \frac{g_{m,i}}{g_{m,k}} p_{m,i}, \quad i \neq k.$$  

(28)

Considering the above equation and the fact that $\sum_{k=1}^{K} p_{m,k} = 1$ we can write:

$$p_{m,k} \left( \sum_{i=1}^{K} \frac{g_{m,k}}{g_{m,i}} \right)^{-1} = 1,$$  

(29)

which leads to following power allocation policy for user $k$ belonging to the cluster $m$:

$$p_{m,k} = \left( \frac{g_{m,k}}{g_{m,i}} \sum_{i=1}^{K} \frac{1}{g_{m,i}} \right)^{-1},$$  

(30)

where $K$ is the number of users associated to the cluster $m$.

VI. NUMERICAL RESULTS

In this section, we investigate the performance of the proposed user-clustering and power allocation schemes via numerical simulations. The main parameters deployed in the numerical simulations are listed in Table I. The radio channel was considered as the product of path loss and Rayleigh fading with zero mean and unit variance. In the considered path-loss model the transmitted signal power decays according to $d_k^{-\alpha}$, where $d_k$ is the distance between BS and the related user $k$ and $\alpha$ is the path-loss exponent. A single cell was considered, with the single BS located at the center of the cell area. The cell radius is set to 500 m and the distance between a user and the BS is $50 < d_k < 500$ m. The circuit power per antenna is assumed $P_c = 30$ dBm, and the basic power consumed at the BS is $P_0 = 40$ dBm. The inefficiency factor of the power amplifier is set to $\alpha = 1.5$. The cluster-heads are assumed to be distributed within a specific distance $r_1$ from the BS, as depicted in Fig. 2.

Fig. 3 depicts the SE, EE and RE of MIMO-NOMA I for different values of $\eta$. As can be seen from the (23), the RE reaches better performance for higher values of $\eta$.

![Figure 2. Illustration of the clustered system model for $K = 12$ and $M = 4$. The BS is considered to have $M = 4$ antennas.](image)

We also assume perfect CSIT. All the simulations are performed considering a single transmission time interval (TTI) with its instantaneous channel gains perfectly estimated. Our numerical results are shown in terms of system SE, EE and RE. Besides, in the following analysis it is assumed the channel gains of the users in each cluster are uncorrelated and identically and independently distributed (i.i.d.). The results obtained through the consideration that the channels of cluster-heads as the equivalent channel vectors are denominated MIMO-NOMA I; while the results associated with the equivalent channel calculation applying eq. (13) are referred as MIMO-NOMA II.

### Table I

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell radius</td>
<td>$R = 500$ [m]</td>
</tr>
<tr>
<td>Minimum distance UE–BS</td>
<td>$d_{\text{min}} = 50$ [m]</td>
</tr>
<tr>
<td>Path loss exponent</td>
<td>$b = 3.5$</td>
</tr>
<tr>
<td>Mobile users</td>
<td>$K = 30$</td>
</tr>
<tr>
<td>BS antennas</td>
<td>$M \in [3; 15]$</td>
</tr>
<tr>
<td>UE antennas</td>
<td>1</td>
</tr>
<tr>
<td>Weighting factor SE-EE</td>
<td>$\eta \in [0.1; 100]$</td>
</tr>
<tr>
<td>Inefficiency of PA</td>
<td>$\alpha = 1.5$</td>
</tr>
<tr>
<td>BS transmit power budget</td>
<td>$P_{T} = 30$ dBm</td>
</tr>
<tr>
<td>Circuit power per antenna</td>
<td>$P_{c} = 30$ dBm</td>
</tr>
<tr>
<td>Basic power consumption BS</td>
<td>$P_{0} = 40$ dBm</td>
</tr>
<tr>
<td>SNR</td>
<td>$\gamma \in [0; 20]$ [dB]</td>
</tr>
<tr>
<td># Monte-Carlo Trials</td>
<td>$T = 500$</td>
</tr>
</tbody>
</table>

Fig. 4 presents the RE of the MIMO-NOMA I system for different number of antennas at the BS and, consequently, for different number of clusters in the cell. By analyzing the difference in the system performance when the number of clusters is changed, one can see that the condition which presents the better performance is when $K = 2$, i.e., in which there are two users per cluster. This is an expected result, since a higher number of antennas at the BS implicates in a higher directivity in the transmission to each cluster, reducing the inter-cluster interference. Also, the resultant smaller number of users in each cluster leads to a reduced intra-cluster interference. Interesting, we have found that instead of having the condition of three users per cluster as the second best performance condition in MIMO-NOMA I, however, we have obtained the condition $M = 5$, which results in a total of six users per cluster, corresponding to the second highest RE results.

Finally, a comparison of the two MIMO-NOMA channel equivalent calculations is presented in Fig. 5 for different system dimen-
sions $M \cdot K$. As one can infer from the figure, and for the system conditions assumed in this paper, the MIMO-NOMA II approach is inferior to MIMO-NOMA I in terms of RE. Also, this difference appears to become smaller for an increase in the number of transmit antennas and users. Among relevant information shown in Fig. 5 one can highlight the reduction in the system performance for a larger number of users ($K = 60$). This indicates there is a limit in the RE performance enhancement achieved by an increase in the value of $K$, after which the interference caused by the additional users implicates in RE degradation. Besides, when the number of BS antennas is not enough to forming suitable clusters with reduced (mitigate) inter-cluster interference, which implies in high number of users per clusters, the RE will be degraded in high SNR.

In this paper, we have analyzed the resource efficiency in a cluster-based MIMO NOMA system through the application of a two-step simple power allocation strategy. Following the user-clustering scheme presented in this work, the cluster head in each downlink MIMO NOMA cluster can almost completely cancel the intra-cluster interference, and, thus, achieves maximum throughput gain in comparison to the other users in the cluster.

Analyzing the system performance in terms of RE, our results demonstrated that the best performance occurs for two users per cluster. However, the following highest RE values are not always attained for the subsequent smaller numbers of users per cluster. Instead, there are conditions where a higher number of users in a cluster was shown to have a superior performance.

A comparison of the two MIMO-NOMA channel equivalent calculations demonstrated that, for the assumed system conditions, the equivalent obtained through the SVD leads to inferior performance results than the direct choice of the cluster-head’s channel. An observation of the system RE for an increase in the number of users also showed that an excessive number of users can result in performance degradation.

**REFERENCES**


[7] Z. Ding, F. Adachi, and H. V. Poor, “The application of mimo to non-orthogonal multiple access,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, Jan 2016.


Appendix B – Sum-Rate Capacity in DL Massive MIMO with Partial CSI and Low-Complexity Linear Precoders
Abstract—The sum-rate capacity achieved by three low-complexity precoders is compared considering a single-cell massive MIMO broadcast channel; these linear precoders include the conventional zero-forcing (ZF) beamforming, regularized channel inversion (RCI) precoding, and a precoding version that does not require a priori knowledge of the channel statistics, which is based on the iterative randomized Kaczmarz algorithm (rKA). The sum-rate capacity vs complexity problem is analyzed in the downlink (DL) with a base-station (BS) deploying massive uniform linear array (ULA) antennas, while mobile terminals (MT) are equipped with single-antenna. We first derive the signal-to-interference-plus-noise (SINR) ratio for the selected precoders and transmission schemes. Numerical results demonstrated that the rKA’s performance-complexity tradeoff is superior compared to conventional ZF and RCI precoding when operating in massive MIMO systems, since it holds a relative robustness against system loading increasing, implies in a considerably reduced number of complex operations and appears to be equally efficient under a simple equal power allocation policy.

Index Terms—Channel Inversion, Kaczmarz algorithm, Massive MIMO, Precoding, Sum-Rate capacity, Zero Forcing

I. INTRODUCTION

In the next generation of communication systems (5G), due to the increasing demand for higher data rates transmission and the exponential growth in the number of users, interference has turn into one of the major limiting factors for performance and capacity of wireless cellular systems. In downlink (DL) transmission of a multiuser MIMO (MU-MIMO) scenario, interference between users is a major source of system errors; hence, schemes able to mitigate it without requiring excessive coordination and control information exchanging are of great interest. These methods are commonly defined in the category of base station (BS) precoding and generally rely on the channel state information (CSI) knowledge.

Beamforming is a widely known technique for interference reduction and directed transmission of energy in the presence of noise and interference. In MIMO systems, some beamforming techniques exploit channel knowledge at the transmitter side to maximize the signal-to-noise ratio (SNR) at the receiver, while others utilize the channel information to transmit in the direction of the eigenvector corresponding to the largest eigenvalue of the channel [1], while canceling the transmission in other directions. Furthermore, beamforming techniques can also be used in the DL of a multi-user system aiming at maximizing the signal-to-interference-plus-noise ratio (SINR) of a particular user [2].

From information theory perspective it is proved that the sum capacity of the MIMO broadcast (spatial multiplexing mode) channel can be achieved through the technique known as dirty-paper-coding (DPC) [3]. However, DPC is a nonlinear precoding scheme and for most practical communication systems it is not feasible due to its very high computational complexity. Due to this reason, researches have focused on quasi-optimal low-complexity approaches. In contrast to the DPC, it has been shown in [4], [5] that sub-optimal linear precoders, such as the matched filtering (MF) precoding, also known as conjugate beamforming, and the zero-forcing (ZF) beamforming can be applied to ensure much lower computational complexity and still providing good performance in terms of achievable sum-rate in the massive MIMO context.

Specifically, the ZF beamforming is a sub-optimal linear precoding or transmit beamforming strategy that is able to cancel the inter-user interference (IUI) simply by pre-multiplying data symbols with the inverse of the channel matrix. However, it has been demonstrated in [6] that the sum-capacity of the ZF does not grow linearly with the number of users, while channel inversion-based precoding strategies can become a serious concern when the channel becomes ill-conditioned. Hence, in order to tackle this problem, a regularization parameter is introduced in the channel inversion and, by that, the corresponding sum-capacity scales linearly with the number of users, but under a slower rate than that achieved by the optimal DPC [6]. This beamforming technique is called regularized channel inversion (RCI) and it does not cancel the inter-user interference completely as the ZF, but it is able to control the amount of interference introduced to each user. This characteristic also contributes to the robustness against noise in comparison to ZF, which may enhance the noise power [7]. Therefore, this regularization parameter should be optimally chosen to maximize some performance indicators, such as the SINR. The optimal regularization parameter when the number of BS antennas is equal to the number of users was derived in [6], while in [8] a generic case was derived by using a large system analysis.

An alluring investigation about precoding techniques for the single-cell DL MU-MIMO systems is carried out in [4]. The authors compared the MF precoding and ZF with respect to spectral-efficiency and radiated energy-efficiency in a single cell scenario. It is showed that, for high spectral-efficiency and low energy-efficiency, ZF outperforms MF, while in the case at low spectral-efficiency and high energy-efficiency the opposite is true. An equivalent result for the uplink (UL) can be found in [9]; the authors have demonstrated that in a low SNR ratio, the simple maximum ratio combining (MRC) receiver outperforms...
the ZF receiver. This can be explained by the fact that, at low power levels, the inter-user interference introduced by the MRC receiver is occasionally less than the noise enhancement caused by the ZF detector; hence, the simple MRC detector becomes a better alternative. This result is analog for the downlink.

In MIMO systems, the BS needs to compute precoding/receiving vectors in order to transmit/detect data to/from each user, which results in a complexity that grows proportionally to the number of antennas $M$ and the number of users $K$. Therefore, the increasing number of antennas in Massive MIMO brings benefits, but they come at the cost of a substantial increase in hardware and computational complexity. Different algorithms have been lately proposed to reduce such complexity. However, the fundamental assumption in these solutions is that the exact statistics of the channel vectors of the users is a priori known. This is not a practical assumption, specially since estimating the channel covariance matrices of the users in a large system regime is also demanding in terms of both computational and storage requirements [12].

In the statistical channel covariance matrix and CSI uncertainty context, a technique for computing the precoder/detector in a massive MIMO system that does not require a priori knowledge of the statistics of the users channel vectors, based on the randomized Kaczmarz algorithm (rKA), was proposed recently in [12]. The KA was initially proposed by Kaczmarz as an iterative technique for solving over-determined (OD) set of linear equations (SLE) [13]. Recently, a randomized version of KA was devised and analyzed for solving consistent OD-SLE [14]. With the increasing popularity of stochastic gradient techniques and machine learning, KA has been revised [15] and applied to other problems such as solving quadratic equations [16]. Moreover, considering under-determined (UD) systems, in [12] an extended version of the rKA algorithm based on [14] is proposed to deal with UD-SLE. In the same work, the rKA methodology is applied to different massive MIMO scenarios, while the performance of both the rKA-based precoder and detector for Massive MIMO is analyzed theoretically. Numerical results in [12] indicate that the rKA precoder/detector schemes achieve a suitable performance while are very competitive from a computational viewpoint.

This work develops a comparative analysis between the RCI, ZF and the rKA precoding schemes in terms of sum-rate capacity under an equal power allocation strategy and realistic massive MIMO channel scenarios, considering 5G target applications, imperfect channel estimation and antenna correlation. The contributions of this work include: a) comparison on the RCI, ZF and rKA-based linear precoders performance-complexity tradeoff; b) analysis of the impact of partial CSI and antenna correlation on the system capacity.

Notation: In the following, boldface lower-case and upper-case characters denote vectors and matrices, respectively. The operators $(\cdot)^H$, $\text{tr} (\cdot)$, $\mathbb{E} [\cdot]$ and $\mathbb{V} [\cdot]$ denote conjugate transpose, trace, expectation and variance, respectively. The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$. A random vector $\mathbf{x} \sim \mathcal{CN} (\mathbf{m}, \Sigma)$ is complex Gaussian distributed with mean vector $\mathbf{m}$ and covariance matrix $\Sigma$. The inner product between two matrices or vectors $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr} (\mathbf{X}^H \mathbf{Y})$. The notations $\| \mathbf{X} \|_F = \sqrt{\mathbb{E} [\mathbf{X}^H \mathbf{X}]}$ and $\| \mathbf{x} \|$ are respectively used for the Frobenius norm of a matrix $\mathbf{X}$ and the $\ell_2$-norm of a vector $\mathbf{x}$.

For two $m \times n$ and $m' \times n$ matrices $\mathbf{X}$ and $\mathbf{Y}$, we denote by $[\mathbf{X} \mathbf{Y}]$ the $(m + m') \times n$ matrix obtained by stacking the rows of $\mathbf{X}$ on top of the rows of $\mathbf{Y}$. $f^t (\cdot)$ is a linear operator that depends on the argument and on the internal randomization of the rKA algorithm until iteration $t$.

II. System Model

In this section, we consider the DL of a single-cell MU-MIMO broadcast channel, where the BS is equipped with $M$ antennas that transmit to $K$ single-antenna user terminals. It is also considered that the large-scale fading between the BS and the receiver user $k$ is denoted by $h_{k}$. Also, it is considered an unbounded path-loss model in which the transmitted signal power decays accordingly to $\frac{1}{r_k^2}$, where $r_k$ is the distance between BS and the related user $k$ and $b$ is the path-loss exponent, which is related to the signal decay behavior in the cell.

With these considerations, the received signal vector $\mathbf{y} \in \mathbb{C}^{K \times 1}$ of a narrow-band communication is given by

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

(1)

where $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmit vector, $\mathbf{A} \in \mathbb{R}^{K \times K} \triangleq \text{diag} (a_1, \ldots, a_K)$ is the path-loss diagonal matrix with $a_k^2 = \frac{1}{r_k^b}$, $\mathbf{H} \in \mathbb{C}^{K \times M}$ the channel matrix and $\mathbf{n} = [n_1 \ldots n_k \ldots n_K]^T \in \mathbb{C}^{K \times 1} \sim \mathcal{CN} (0, \sigma^2_n \mathbf{I}_K)$ the i.i.d Additive White Gaussian (AWG) noise vector. The transmit signal vector $\mathbf{x}$ is obtained from the product of a linear precoding $\mathbf{G} \in \mathbb{C}^{M \times K} \triangleq [\mathbf{g}_1 \ldots \mathbf{g}_K]$ with the symbol vector $\mathbf{s} = [s_1 \ s_2 \ldots s_K]^T \in \mathbb{C}^{K \times 1}$, which is normalized in power, i.e., $\mathbb{E} [\mathbf{s} \mathbf{s}^H] = \mathbf{I}_K$. Hence, the linear precode $\mathbf{G}$ and the power matrix $\mathbf{P} \in \mathbb{R}^{K \times K} = \text{diag} (p_1, \ldots, p_K)$ allocated for each of $K$ users define the transmit vector:

$$
\mathbf{x} = \mathbf{GP}^{1/2} \mathbf{s} \equiv \sum_{k=1}^{K} \sqrt{p_k} \mathbf{g}_k s_k.
$$

(2)

Indeed, the transmit vector $\mathbf{x}$ can also be expressed as a linear combination of the independent user symbols $s_k$, where $\mathbf{g}_k \in \mathbb{C}^{M \times 1}$ and $p_k \geq 0$ are the precoding vector and the signal power of the $k$-th user, respectively. We are working under the assumption of perfect channel state information at the transmitter (pCSIT), as well as under channel error estimates (eCSIT).

Moreover, the precoding vectors are normalized to satisfy the average power constraint, thus:

$$
\mathbb{E} \left[ \| \mathbf{x} \|^2 \right] = \text{tr} (\mathbf{P} \mathbf{G}^H \mathbf{G}) \leq P_T,
$$

(3)

where $P_T \geq 0$ is the total available transmit power at the BS. Hence, the SNR at the transmitter side is denoted by $\gamma = \frac{P_T}{\sigma^2_n}$. 
1) MIMO Correlation Channels: Correlated MIMO channel scenarios are being considered, as in [18, eq.(6)-(7)]:
\[ h_k = \hat{h}_k \Phi^{1/2}, \]
where \( h_k \in \mathbb{C}^{1 \times M} \) is the \( k \)-th row of the channel matrix \( \mathbf{H} \), and the covariance channel matrix is denoted by
\[ \Phi = E[\{h_k^H h_k^\dagger\}], \]
where \( \hat{h}_k \sim \mathcal{CN}(0_{1 \times M}, I_M) \) is the uncorrelated small-scale channel fading vector for the user \( k \), and \( M \) is the number of BS antennas. This is a general channel model family for massive MIMO channels, which can represent the following channel scenarios:
a) i.i.d. Rayleigh case (rich scattering environment) with \( \Phi = I_M \).
b) spatial correlated channels under uniform linear antenna (ULA) array, where the correlation matrix is defined by:
\[ [\Phi]_{i,j} = e^{j\pi c i j}, \quad \text{with the correlation parameter } c \in [0, 1]. \]
c) planar antenna array (UPA) correlation model. A multidimensional array correlation structure is constructed for the UPA based on a Kronecker product of two ULA correlation matrices as in [19]. We assume an approximation in which the correlation between elements along \( x \) coordinate does not depend on \( y \) and is given by matrix \( \Phi_x \), and the correlation along \( y \) coordinate does not depend on \( x \) and is given by matrix \( \Phi_y \). The following Kronecker-type approximation of the UPA correlation matrix is proposed by [19]:
\[ \Phi \approx \Phi_x \otimes \Phi_y = \begin{bmatrix} \phi_{x1,1} \mathbf{R}_y & \cdots & \phi_{x1,n_x} \mathbf{R}_y \\ \vdots & \ddots & \vdots \\ \phi_{x n_x,1} \mathbf{R}_y & \cdots & \phi_{x n_x,n_y} \mathbf{R}_y \end{bmatrix}, \]
where \( \otimes \) denotes the Kronecker product. UPA correlation matrix at the receiver \( \Phi \) is the Kronecker product of two ULA correlation matrices \( \Phi_x \) and \( \Phi_y \), which are Toeplitz. Therefore, even though \( \Phi \) may not be a Toeplitz matrix, its approximation (7) has a Toeplitz per block structure.

According to [20], the approximation model for UPA correlation matrix in (7) is reasonably accurate, allowing the usage of the well-developed theory of Toeplitz matrices for the analysis of multidimensional antenna arrays.

2) Noisy Version for the MIMO Channel Matrix: Besides, the channel matrix estimation available at BS \( \mathbf{Q} \) in each coherence time period \( \Delta T_c \) is not perfect, i.e., \( \mathbf{Q} \neq \mathbf{H} \). Indeed, estimated channel matrix \( \mathbf{Q} \) is a noisy correlated version of the true channel matrix \( \mathbf{H} \), possibly modelled as:
\[ \mathbf{Q} = \sqrt{1-\tau^2} \mathbf{H} + \tau \mathbf{N}, \quad \text{where } [\mathbf{N}]_{k,m} \sim \mathcal{CN}(0, 1) \]
where \( \tau \in [0; 1] \) is the estimation quality parameter, with \( \tau = 0 \) yielding a perfect channel estimation, \( \mathbf{Q} = \mathbf{H} \) [12]. With the previous considerations, the received DL vector \( \mathbf{y} \) from (1) can be rewritten as:
\[ \mathbf{y} = \mathbf{A H x} + \mathbf{n} = \mathbf{A H G P}^{1/2} \mathbf{s} + \mathbf{n} \]
and the received DL signal \( y_k \) at the \( k \)-th UE is given as:
\[ y_k = a_k \sqrt{p_k} h_k g_k s_k + \sum_{j=1, j \neq k}^{K} a_j \sqrt{p_j} h_j g_j s_j + n_k. \]
As a result, the SINR per user is defined as [10]:
\[ \text{SINR}_k = \frac{a_k^2 p_k |h_k g_k|^2}{\sum_{j=1, j \neq k}^{K} a_j^2 p_j |h_j g_j|^2 + \sigma_n^2}, \quad \forall k = 1, \ldots, K \]
and the normalized achievable rate of user \( k \), in terms of [bits/s/Hz] is given as:
\[ R_k = \log_2 (1 + \text{SINR}_k). \]
Moreover, the ergodic sum-rate capacity is given by
\[ R_S = \mathbb{E} \left[ \sum_{k=1}^{K} \log_2 (1 + \text{SINR}_k) \right], \]
where the expectation is taken over a large number of channel realizations \( \mathbf{h}_k \), \( \forall k = 1, \ldots, K \).

III. LINEAR BEAMFORMING SCHEMES

In this section we derive the sum-rate capacity for the conventional ZF and RCI linear precoders, as well as for an iterative low-complexity precoder based on the rKA [12].

A. Zero-Forcing Beamforming

The ZF precoding, also referred as channel inversion (CI) precoding, eliminates all the inter-user interference by performing an inversion of the channel matrix \( \mathbf{H} \) at the transmitter side [21]. ZF beamforming is largely applied to MU-MIMO networks with single antenna users [4], [10]. Due to its simplicity on designing beamforming vectors \( \mathbf{g}_k \), it allows the users to receive data free of interference, which can be attained thanks to the orthogonality imposed by the beamforming vectors for different users. The precoding matrix for the ZF is given by [21]
\[ \mathbf{G}_{ZF} = \alpha \mathbf{H} \mathbf{H}^H \mathbf{H}^{-1} = \alpha \left( \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{H}^H, \]
where the normalization constant \( \alpha \) is chosen to satisfy the power constraint in (3) and it is built only upon the channel realization \( \mathbf{H} \), being determined by
\[ \alpha = \sqrt{\frac{P_t}{\text{tr} \left( \mathbf{P} \left( \mathbf{H H}^H \right)^{-1} \right)}}. \]

The received vector can be represented as
\[ \mathbf{y} = \alpha \mathbf{A H H}^H \mathbf{H}^{-1} \mathbf{P}^{1/2} \mathbf{s} + \mathbf{n} = \alpha \mathbf{A P}^{1/2} \mathbf{s} + \mathbf{n}. \]
With these considerations, the SINR of user \( k \) under ZF precoding is given as
\[ \text{SINR}_{ZF}^k = \frac{\alpha^2 a_k^2 p_k}{\sigma_n^2}, \quad \text{with } K \leq M. \]
By using the ZF beamforming, the precoding vector is constructed to eliminate the interference that a particular user may cause to others, i.e., IUI.

Notice that the ZF has a limited number of users, which is bounded by the number of BS antennas, $K \leq M$. If the number of single-antenna users, $K$, increase beyond $M$, the IUI is much stronger in the system and the SINR described in (17) does not hold. For cases where $K > M$, the received signal vector $y$ and signal $y_k$ are given, respectively, by:

$$y = \alpha A H (H^H H)^{-1} H^H P^{1/2} s + n$$

$$y_k = \alpha_k \sqrt{\mu_k} h_k (H^H H)^{-1} h_k^H s_k +$$

$$+ \alpha \sum_{j=1,j \neq k}^K a_k \sqrt{\mu_j} h_k (H^H H)^{-1} h_j^H s_j + n_k,$$

leading to a SINR for the $k$th user given by

$$\text{SINR}_{k}^{ZF} = \left. \frac{\alpha^2 \alpha_k^2 \mu_k |h_k (H^H H)^{-1} h_k^H|^2}{\alpha^2 \sum_{j=1,j \neq k}^K \alpha_k^2 \mu_j |h_k (H^H H)^{-1} h_j^H|^2 + \sigma_n^2} \right|_{k}.$$

with the term $\alpha^2 \sum_{j=1,j \neq k}^K \alpha_k^2 \mu_j |h_k (H^H H)^{-1} h_j^H|^2$ representing the IUI.

Since the SINR$_{k}^{ZF}$ is proportional to $\alpha^2$, a rank deficiency on the channel correlation matrix $HH^H$ will lead to a penalty in $\alpha^2$ and, consequently, in the SINR. This motivates the addition of a regularization parameter, as discussed in the sequel.

### B. Regularized Channel Inversion

The RCI precoding can be faced as a generalization of the ZF precoding, in which a regularization parameter is included in the pseudo-inverse matrix. Basically, to compensate the possibility of an ill-conditioned channel matrix $H$, for instance, in case of spatial channel correlation, the regularization term $\xi$ is added within the pseudo-inverse of the ZF precoding matrix, in Eq. (14). Hence, considering a system equipped with the RCI precoder, the precoding matrix solution is given by [11], [22], [23]:

$$G_{\text{RCI}} = \alpha H (H H^H + \xi I_M)^{-1},$$

or equivalently as

$$G_{\text{RCI}} = \alpha (H^H H + \xi I_M)^{-1} H^H,$$

where the normalizing constant $\alpha$ is chosen to satisfy the power constraint (3), and $\xi > 0$ is the regularization parameter.

As stated in [21], by assuming independence between the data symbols, similarly to (15), the normalization constant for the RCI precoding is expressed as

$$\alpha = \left[ \frac{P_t}{\text{tr} \left( P H (H^H H + \xi I_M)^{-2} H^H \right)} \right].$$

where now the normalization factor $\alpha$ depends on the channel realization $H$, as well as the regularization factor $\xi$.

Using the RCI precoder (21), the received vector $y$ and signal $y_k$ for each user can be respectively expressed as

$$y = \alpha A H (H^H H + \xi I_M)^{-1} H^H P^{1/2} s + n$$

$$y_k = \alpha \alpha_k \sqrt{\mu_k} h_k (H^H H + \xi I_M)^{-1} h_k^H s_k +$$

$$+ \sum_{j=1,j \neq k}^K \alpha_k \sqrt{\mu_j} h_k (H^H H + \xi I_M)^{-1} h_j^H s_j + n_k,$$

where the first term in the right side of (24) is the desired signal for each user $k$, the second term is the interference introduced by the other users and the last one is the received thermal noise.

In order to avoid excessive complexity, we advocate the use of single-user detection at the DL receiver side. Hence, the SINR for each user is expressed as follows [11][24]

$$\text{SINR}_{k}^{\text{RCI}} = \left. \frac{\alpha^2 \mu_k |h_k (H^H H + \xi I_M)^{-1} h_k^H|^2}{\alpha^2 \sum_{j=1,j \neq k}^K \mu_j |h_k (H^H H + \xi I_M)^{-1} h_j^H|^2 + \sigma_n^2} \right|_{k}.$$

Notice that when $\xi \rightarrow 0$, the RCI precoder converges to the ZF precoder. This relationship is direct, since the term $(H^H H + \xi I_M)^{-1} H^H$ will tend to $(H^H H)^{-1} H^H$ as $\xi$ tends to zero.

### C. rKA-based Precoding

The iterative rKA precoder, proposed in [12] as an extension of the KA [13] and the rKA [14], does not require a priori knowledge of the statistics of the users channel vectors in a massive MIMO system. In order to explore the potential of rKA precoding, including complexity reduction, we first define a possibly quantized and noisy estimate of the true channel matrix $H$ available at the BS by $Q \in C^{K \times M}$ as defined in (8). For this purpose, all computation performed at the BS is then assumed to be based on this estimate matrix $Q$. We consider the RCI as the linear precoding scheme, with the transmit vector calculated as

$$x = G s = \alpha Q H^H (Q Q^H + \xi I_M)^{-1} P^{1/2} s.$$

Finding the suitable $M$-dimensional transmit vector to be transmitted to the users in the DL direction can be interpreted as finding the solution $w^*$ of a SLE of the form $Aw = b$, where $A \in C^{n \times n}$ and the vectors $w \in C^{n \times 1}$ and $b \in C^{n \times 1}$ are, respectively, associated with the estimate channel matrix $Q$, the transmit vector $x$ and the symbol vector $s$. Herein, the optimal solution is found via the rKA iterative procedure presented in Algorithm 1.

The rKA algorithm works by selecting one of the rows, say row number $r(t)$, of the matrix $A$ at each iteration $t$. This selection is random with a specific probability. Then it calculates an estimate $w^t$ of the optimal solution $w^*$. If $w^t$ satisfies the equation expressed by the conjugate of the $r(t)$-th row of $A$ denoted by $a_{r(t)}$, i.e., if $\langle a_{r(t)}, w^t \rangle = b_{r(t)}$, then

1 Since rKA-based precoding does not depend on the matrix inversion.
\( \mathbf{w}^t \) is hold. Otherwise, the algorithm updates \( \mathbf{w}^t \) along \( \mathbf{a}_{r(t)} \) to make the \( r(t) \)-th equation consistent. This update, made at each iteration \( t \), can be written as
\[
\mathbf{w}^{t+1} = \mathbf{w}^t + \frac{b_{r(t)} - \langle \mathbf{a}_{r(t)}, \mathbf{w}^t \rangle}{\| \mathbf{a}_{r(t)} \|^2} \mathbf{a}_{r(t)}, \tag{27}
\]

At each iteration \( t \) the algorithm selects a row of \( \mathbf{A} \) randomly and independently from the previous iterations, according to a probability distribution \( \mathbf{p} = (p_1, \ldots, p_m)^T \), where \( p_i \in [0, 1] \) is the probability of selecting the \( i \)-th row of \( \mathbf{A} \) and \( \sum_{i=1}^m p_i = 1 \).

The main parameter that controls the convergence speed of the rKA is the average gain \( \kappa_\chi(\mathbf{A}, \mathbf{p}) \) of the matrix \( \mathbf{A} \) over the subspace \( \chi \in \mathbb{C}^n \) generated by the conjugate of the rows of \( \mathbf{A} \), defined in [12, Definition 1], which depends both on the matrix \( \mathbf{A} \) and the probability distribution \( \mathbf{p} \). Indeed, the work in [12] also demonstrates \( \kappa_\chi(\mathbf{A}, \mathbf{p}) = \frac{1}{\min(m,n)} \), and that the closer \( \kappa_\chi(\mathbf{A}, \mathbf{p}) \) is to \( \frac{1}{\min(m,n)} \), the faster rKA converges across the iterations. Thus, the optimal probability distribution vector \( \mathbf{p} \) can be found through finding the distribution \( \mathbf{p} \) that maximizes \( \kappa_\chi(\mathbf{A}, \mathbf{p}) \). However, although this problem can be solved via convex optimization techniques, this incurs in the same order of complexity as directly computing the precoding vectors, thus being not suitable for the application analyzed in this section, since we are looking for techniques with substantial complexity reduction.

Instead of using the optimal probability distribution mentioned above, we use the suboptimal distribution \( \mathbf{p} \) proposed by [14], where \( p_i = \frac{\| \mathbf{a}_i \|^2}{\| \mathbf{A} \|^2} \) and the probability of selecting the \( i \)-th row scales proportionally to its \( \ell_2 \)-norm \( \| \mathbf{a}_i \|^2 \). Computing this probability distribution incurs in a complexity of \( \mathcal{O}(mn) \), which is much lower than directly solving the maximization problem to find the optimal \( \mathbf{p} \).

It is demonstrated in [12] that, for a fixed loading factor \( \beta = \frac{K}{M} \), the matrix gain \( \kappa_\chi(\mathbf{A}, \mathbf{p}) \) gives an approximation of the best matrix gain \( \kappa_\chi(\mathbf{A}, \mathbf{p}^\star) \) up to a multiplicative factor. This also leads to the conclusion that the suboptimal rKA converges slower than the optimally-tuned rKA up to the same multiplicative factor.

Notice that the KAs proposed in [13] and [14] are applicable to consistent overdetermined (OD) cases. For the massive MIMO DL application, however, we need variants of KA for OD scenarios, where the equations are almost always inconsistent. The work in [12] proposes a new KA variant which uses a step prior to the rKA to remove the inconsistency in the linear equations.

Based on (26), we deploy the new version of KA inspired in [12] that is able to find the signal \( \mathbf{w} = (\mathbf{QQ}^H + \xi \mathbf{I}_K)^{-1} \mathbf{s} \), considering \( \mathbf{P} = \mathbf{I}_K \) since uniform power allocation is assumed, from which one can obtain \( \mathbf{x} = \alpha \mathbf{Q}^H \mathbf{w} \) and solve the precoding massive MIMO problem in a simple, low-complexity but effective way.

Considering \( \mathbf{s} \in \mathbb{C}^{K \times 1} \) the symbol sequence to be sent from the BS to the mobile users (DL), and \( \mathbf{x} = \alpha \mathbf{Q}^H (\mathbf{QQ}^H + \xi \mathbf{I}_K)^{-1} \mathbf{s} \) being the \( M \)-dim precoded transmitted signal, we can apply the rKA in order to find the \( K \)-dim signal \( \mathbf{w} = (\mathbf{QQ}^H + \xi \mathbf{I}_K)^{-1} \mathbf{s} \), from which we can write \( \mathbf{x} = \alpha \mathbf{Q}^H \mathbf{w} \).

According to [12, Proposition 3], we can obtain \( \mathbf{w} = (\mathbf{QQ}^H + \xi \mathbf{I}_K)^{-1} \mathbf{s} \) as the optimal solution of the minimization problem \( \| \mathbf{Bx} - \mathbf{b} \|^2 \), where \( \mathbf{B} = [\mathbf{Q}^H, \sqrt{\xi} \mathbf{I}_K] \) is an \((M+K) \times K\) matrix and \( \mathbf{b} = [0; \sqrt{\xi} \mathbf{s}] \) is an \((M+K)\)-dim vector. This leads to the OD SLE:
\[
\mathbf{Bw} = \mathbf{b}, \tag{28}
\]

that must be solved for the vector \( \mathbf{w} \), from which we can obtain the desired precoded signal \( \mathbf{x} = \alpha \mathbf{Q}^H \mathbf{w} \) at the BS massive antennas side. However, the SLE in (28) is inconsistent unless \( s = 0 \). Thus, in order to remove this inconsistency, we define \( \mathbf{z} = \mathbf{Bw} \) and solve the UD SLE defined as
\[
\mathbf{B}^\perp \mathbf{z} = \mathbf{B}^\perp \mathbf{b} = \sqrt{\xi} \frac{\mathbf{s}}{\sqrt{\xi}} = \mathbf{s} \tag{29}
\]
for \( \mathbf{z} \) through the application of the rKA to (29), from which we obtain the estimate \( \hat{\mathbf{z}} \). Then, we use \( \hat{\mathbf{z}} \) to solve the OD but consistent SLE
\[
\mathbf{Bw} = \hat{\mathbf{z}}, \tag{30}
\]
which leads us to find \( \mathbf{w} \). Assuming \( \mathbf{B} \) as defined previously, with \( \mathbf{B}_{r(t)} \) being the conjugate of the \( r(t) \)-th row of \( \mathbf{B}^\perp \) and denoting \( \mathbf{z} = [\mathbf{u}; \sqrt{\xi} \mathbf{v}] \) for \( \mathbf{u} \in \mathbb{C}^{M \times 1}, \mathbf{v} \in \mathbb{C}^{K \times 1} \), the suboptimal probability that the row \( r(t) \) be selected in the rKA for solving (29) is obtained through:
\[
\mathbf{r}_{r(t)} = \frac{\| \mathbf{B}_{r(t)} \|^2}{\| \mathbf{B}^\perp \|^2} = \frac{\| \mathbf{q}_{r(t)} \|^2 + \xi}{\| \mathbf{Q}^H \|^2 + K \xi}, \tag{31}
\]
for \( \mathbf{q}_{r(t)} \) as the conjugate of \( r(t) \)-th row of \( \mathbf{Q} \). The update step of the rKA (27) applied for solving (29) can be written as:
\[
\mathbf{z}^{t+1} = \mathbf{z}^t + \frac{\mathbf{s}_r(t) - \langle \mathbf{B}_{r(t)}, \mathbf{z}^t \rangle}{\| \mathbf{B}_{r(t)} \|^2} \mathbf{B}_{r(t)}. \tag{32}
\]
Moreover, the residual term \( \mathbf{v}^j \) can be rewritten as:
\[
\mathbf{v}^j = \frac{\mathbf{s}_r(t) - \langle \mathbf{q}_{r(t)}, \mathbf{u}^j \rangle}{\| \mathbf{q}_{r(t)} \|^2} - \xi \mathbf{v}_{r(t)}^j, \tag{33}
\]
which leads to the components \( \mathbf{u} \) and \( \mathbf{v} \) of vector \( \mathbf{z} \) being given respectively as:
\[
\mathbf{u}^{t+1} = \mathbf{u}^t + \mathbf{v}^j \mathbf{q}_{r(t)}, \tag{34}
\]
\[
\mathbf{v}_{r(t)}^{t+1} = \mathbf{v}_{r(t)}^t + \mathbf{v}^j, \quad \text{and} \quad \mathbf{v}_{j}^{t+1} = \mathbf{v}_{j}^t \quad \text{for} \quad j \neq r(t). \tag{35}
\]
Finally, the desired precoding vector \( \mathbf{w} \) in (30) can be obtained as the rKA estimate of \( \mathbf{v} \). The entire process is summarized by the pseudocode in Algorithm 1.

Algorithm 1 is used to directly compute a suitable transmit vector for the DL of the massive MIMO system. However, computing the DL precoding matrix from the available estimation of the channel state can be useful when the channel coherence block is large, which results in the precoding matrix corresponding to each block being computed only once inside the block and in reducing the computational complexity [12].
Algorithm 1 Kaczmarz Algorithm for DL (Precoding)

1: Input: $Q \in \mathbb{C}^{K \times M}$, $s \in \mathbb{C}^{K \times 1}$ and RCI parameter $\xi \geq 0$.
2: Define $u_t$ and $v_t$ and initialize $u^0 = 0$, $v^0 = 0$.
3: Apply KA to Eq. (29) through:
4: for $t = 0, 1, \ldots, T - 1$ do
5: Pick a row $r(t)$ of $Q$ with prob.: $p_t(r) = \frac{||q_{r(t)}||^2 + \xi}{||Q||^2_F + \xi}$
6: Compute the residual $y^t$ as in Eq. (33).
7: Update $u$, Eq. (34).
8: Update $v$, Eq. (35).
9: end for
10: Output: Set $w = v^{T-1}$, Eq. (30).
11: Output: Set $x = \alpha Q^H w$.

**1) Computing Precoding Matrix on a Specific Coherence Block using rKA:** With the goal of generalizing the rKA procedure aiming to compute the precoding matrix directly, we focus on a specific coherence block. We then use Algorithm 1 and run $K$ KA procedures in parallel, where the input for the $i$-th KA is $s_i = e_i \in \mathbb{C}^{K \times 1}$, for $e_i$ denoting the $i$-th canonical basis, with “1” in its $i$-th component and “0” elsewhere. The $K$ parallel KA processes may share the same random row index at each iteration or may use their independent randomness. After a suitable convergence, the output of the $i$-th KA using $s_i = e_i$ is denoted by $w_i \in \mathbb{C}^{M \times 1}$, and the corresponding precoding vector would be $x_i = \alpha Q^H w_i$. Due to the convergence of KA, $x_i$ should be approximately equal to

$$x_i \approx \alpha Q^H (Q Q^H + \xi I_K)^{-1} e_i,$$

(36)

which corresponds to the $i$-th column of the precoding matrix $\alpha Q^H (Q Q^H + \xi I_K)^{-1}$. Since $Q$ is assumed to have a full column rank, one can conclude that $w_i$, after a suitable convergence, corresponds to the $i$-th column of $(Q Q^H + \xi I_K)^{-1}$.

Hence, $(Q Q^H + \xi I_K)^{-1}$ can be approximated by the matrix $W = \begin{bmatrix} w_1 & \ldots & w_i & \ldots & w_K \end{bmatrix} \in \mathbb{C}^{K \times K}$ consisting of the outputs of the $K$ parallel KAs with inputs $s_i = e_i$, $i \in \{1, 2, \ldots, K\}$. It also implies the precoding matrix can be approximated as $G = \alpha Q^H W$. As discussed in Section V-E, in terms of computational complexity it is beneficial to compute the precoding vector following the multiplication order:

$$x = (\alpha Q^H (W s)),$$

(37)

where the parenthesis indicate the order of multiplication.

With the purpose of summarizing the precoders presented in this work and to provide a better comparison between them, the precoding matrix, SINR and computational complexity in terms of complex operations of each precoder structure are organized in Table 1.

**IV. LARGE SPECTRUM SINR ANALYSIS**

In this section, a large limit analysis for the SINR and the achievable ergodic rate is provided. Specifically, the SINR of the RCI and the ZF precoding schemes are analyzed under a large system limit where both the number of users $K$ and the number of transmit antennas $M$ approach infinity, with a fixed ratio $\beta = \frac{K}{M} < 1$. Large system limit analysis expressions for the asymptotic sum-rate capacity were previously studied in [21], [11] and [24]. Capacity bounds for the rKA precoder are provided based on [12] and [25]. Herein, the analysis is applied to the DL of a massive MIMO system equipped with an $M$ antennas ULA at the BS, and uniform power allocation is considered.

**A. Sum-Rate under Transmit-Side Channel Correlation**

In Appendix C the SINR expressions for ZF and RCI precoders operating under large system regime are derived aiming to determine the sum-rate capacity of such linear precoding. Hence, the achievable sum-rate capacity is

$$\mathcal{R}_\Sigma^\infty = \sum_{k=1}^{K} \log_2 \left( 1 + \frac{\left|T_{dl, k}^i\right|^2}{\sum_{k' \neq k} \left|T_{dl, k'}^i\right|^2 + \sigma_k^2} \right),$$

(38)

where $\mathcal{R}_\Sigma^\infty$ is the SINR for the $k$th user in RCI and ZF precoding defined by eq. (64) and (65), respectively. Notice that the asymptotic SINR expressions for both linear precoders are different for each user and depend on the pathloss $\alpha_k$.

**B. Capacity Bounds for rKA Precoder**

The lower and upper bounds for the achievable ergodic rate when rKA precoding is deployed are presented in Appendix A, based on [12], and originally developed in [25].

For the DL massive MIMO scenario, an upper bound on the achievable ergodic rate when the rKA precoding is run for a specific number of iterations $t$, can be obtained assuming time-varying interference matrix, by treating the interference as noise and by coding across several coherence blocks, being evaluated by:

$$\mathcal{R}_k = \mathbb{E} \left[ \log_2 \left( 1 + \frac{\left|T_{dl, k}^i\right|^2}{\sum_{k' \neq k} \left|T_{dl, k'}^i\right|^2 + \sigma_k^2} \right) \right],$$

(39)

where

$$T_{dl} = H g^i(Q) = H G \in \mathbb{C}^{K \times K}.$$

(40)

For simplicity, we dropped the explicit dependence of $T_{dl}$ on the iteration $t$. The expectation in (39) is taken over all the randomness of the channel state $H$ and the precoding matrix $g^i(Q)$, including the randomness due to KA. This yields the upper bound $\mathcal{R} = \sum_{k=1}^{K} \mathcal{R}_k$ on the achievable sum rate.

Notice that the upper bound $\mathcal{R}$ was obtained under the condition that the true channel state $H$ is available at the BS, such that the channel interference matrix $T_{dl}$ is fully known to the BS. This yields an upper bound on the achievable ergodic rate in our scenario where only a noisy version of $H$ is available at the BS, thus, the channel interference matrix $T_{dl}$ is not perfectly known at the BS (since $H$ is not known).

For the lower bound on the ergodic rate, and following the derivation from [25], we can write:

$$\mathcal{R}_\lambda = \log_2 \left( 1 + \frac{\mathbb{E}[T_{dl, k}^i]^2}{\sqrt{\mathbb{E}[T_{dl, k}^i]^2} + \sum_{k' \neq k} \mathbb{E}[T_{dl, k'}^i]^2} + \mathbb{E}[\sigma_k^2] \right).$$

(41)

Varying across coherence time intervals due to channel randomness, as well as varying across multiple slots inside a coherence block due to internal randomness of KA.
### Table I
PRECODING STRUCTURE, MATRIX AND SINR RELATIONS.

<table>
<thead>
<tr>
<th>Precoding</th>
<th>Precoding Matrix, <strong>G</strong></th>
<th>SINR&lt;sub&gt;κ&lt;/sub&gt;</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZF (K ≤ M)</td>
<td>α&lt;sup&gt;H&lt;/sup&gt; (HH&lt;sup&gt;H&lt;/sup&gt;)&lt;sup&gt;−1&lt;/sup&gt;</td>
<td>(\frac{\alpha^2 a_k^2 p_k}{\sigma_n^2})</td>
<td>(O(K^2 M) + O(K^2) + O(M K)) [26]</td>
</tr>
<tr>
<td>ZF (K &gt; M)</td>
<td>(\alpha (H^H H)^{−1} H^H)</td>
<td>(\frac{\alpha^2 a_k^2 p_k}{\sigma_n^2} \left</td>
<td>h_k (H^H H)^{−1} h_j^H \right</td>
</tr>
<tr>
<td>RCI</td>
<td>(\alpha (H^H H + \xi I_M)^{−1} H^H)</td>
<td>(\frac{\alpha^2 a_k^2 p_k}{\sigma_n^2} \left</td>
<td>h_k (H^H H + \xi I_M)^{−1} h_j^H \right</td>
</tr>
<tr>
<td>Iterative rKA</td>
<td>(Q^H (QQ^H + \xi I_K)^{−1} Q)</td>
<td>(\frac{\sum_{k',j,k}</td>
<td>T_{k,k'}^{m}</td>
</tr>
</tbody>
</table>

which leads to the lower bound \(R = \sum_{k=1}^{K} R_k\) on the ergodic rate. Hence, the DL system ergodic rate with rKA precoding is bounded by:

\[
R \in [\hat{R}, \bar{R}] = \left[ \sum_{k=1}^{K} \mathbb{E}[R_k], \sum_{k=1}^{K} \mathbb{E}[\bar{R}_k] \right]. \tag{42}
\]

In order to calculate \(\mathcal{C}(Q)\) for each iteration \(t\) explicitly, we can apply (54) in Appendix B. As an alternative, one can apply the technique described for computing the precoding matrix directly. As an example, in the DL scenario, we can run \(K\) rKAs in parallel with the corresponding inputs \(s_i = e_i, i = 1, \ldots, K\) for calculating the precoding matrix \(G\). Using these concepts, at each instance of the Monte-Carlo (MC) simulation it is possible to compute \(\mathcal{C}(Q)\), producing independent realizations of the random variables needed for computing the upper and lower bounds in (39) and (41), respectively.

### V. NUMERICAL RESULTS

This section provides a comparison on the sum-rate capacity of the ZF beamforming, the RCI and rKA precoders for different system arrangements, including distinct number of antennas at the BS and number of users in the cell and perfect and imperfect CSI conditions. Besides, the rKA precursor convergence is evaluated taking into consideration the channel estimation and correlation effects. We evaluate and compare the proposed precoding schemes under equal power allocation (EPA). Moreover, the analytical large scale limit SINR and sum-rate capacity results are analyzed and compared with the numerical simulation results aiming to corroborate its precoders behavior considering different system configurations.

#### A. Simulation Setup

The main simulation parameters are presented in Table II. We have considered an uniformly distributed random position inside the cell for each user. As an example, Fig. 1 depicts an \(M \times K = 64 \times 64\) system in a \(R = 500\) m macro-cell with uniformly distributed users. We considered an urban cellular radio environment, which leads to a path-loss exponent \(b = 3.5\). For correlated channels we have deployed a simple yet effective ULA spatial correlation model of eq. (6) for low and medium spatial correlation, i.e., adopting \(\epsilon = 0.2\) and \(\epsilon = 0.5\), respectively. For channel error estimation, we deployed the model given in eq. (8).

#### B. rKA Precoder – Typical Operation

Fig. 2 depicts the upper and lower capacity bounds (UB and LB, respectively) for the rKA precoder, and give us an idea on how many iterations \(T\) are necessary for the rKA precoder attain the UB and LB, respectively. Hence, considering typical system and channel parameters, including loading factor \(\beta = 0.5\), spatial ULA correlation \(\epsilon = 0.5\), estimation quality \(\tau = 0.1\) and EPA, the number of necessary iterations for the rKA precoder to achieve full convergence
Table II
SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell radius</td>
<td>$R = 500$ [m]</td>
</tr>
<tr>
<td># BS antennas</td>
<td>$M \in [64, 128]$</td>
</tr>
<tr>
<td># UE antennas</td>
<td>$K \in [32, 64]$</td>
</tr>
<tr>
<td># mobile users</td>
<td>$\beta = 0.5$</td>
</tr>
<tr>
<td>Loading factor</td>
<td>$\gamma \in [0; 0.2; 0.5]$</td>
</tr>
<tr>
<td>Regularization parameter</td>
<td>$\xi \in [0; 0.95]$</td>
</tr>
<tr>
<td>Spatial ULA correlation</td>
<td>$\tau \in [0; 50]$</td>
</tr>
<tr>
<td>Estimation quality parameter, $c$</td>
<td>$\tau \in [0; 0.95]$</td>
</tr>
<tr>
<td>Path loss exponent</td>
<td>$b = 3.5$</td>
</tr>
<tr>
<td>Minimum distance UE-BS</td>
<td>$d_{\text{min}} = 50$ m</td>
</tr>
<tr>
<td>SNR</td>
<td>$\gamma \in [0; 20]$ [dB]</td>
</tr>
<tr>
<td>#rKA iterations</td>
<td>$T \in [10, 900]$</td>
</tr>
<tr>
<td># trials</td>
<td>$T = 500$</td>
</tr>
</tbody>
</table>

of the UB and LB sum-rate capacities is of about $T = 400$ iterations.

\[
\text{Figure 2. Capacity Upper and Lower Bounds of the rKA precoder for } \beta = 0, \tau = 0.1, c = 0.5 \text{ and } T = [100, 250, 300, 450, 400]; \text{ equal power allocation.}
\]

1) Numerical and Analytical Capacity for the ZF, RCI and rKA: Numerical and analytical ergodic capacity calculations for the ZF, RCI and rKA precoding schemes under perfect CSI knowledge at the transmitter side are depicted in Fig. 3. For such analysis we have considered various cell-loading ($\beta$) and SNR ($\gamma$) conditions. The numerical results were analyzed considering $M \in [64, 128]$. The analytical capacity results for ZF and RCI were included through the expressions shown in the large system analysis developed in Section IV, and were calculated using normalized equal power values.

Fig. 3 illustrates the numerical sum-rate results of the three analyzed precoding schemes, evaluated for different system dimensions and correlation parameters, i.e., non correlation ($c = 0$), and medium channel correlation ($c = 0.5$). It can be observed that the large system analytical analysis fit well with the ones evaluated by Monte-Carlo simulations. It can also be seen that, when the number of iterations applied is $T_{\text{full convg}}$, rKA has a similar performance to that of RCI for all SNR conditions and for different system dimensions.

The ZF precoding, in comparison, presents a more degraded performance than the other two for low SNR values, but reaches similar performance for higher SNR conditions. Also, the difference between the sum-rate of ZF and RCI/rKA becomes larger for higher system dimensions.

C. rKA Convergence

In order to analyze the convergence speed of rKA, we compute the minimum gain $\kappa_{\chi}(B^H)$ of matrix $B^H$ along the subspace $\chi$ spanned by the conjugate of the rows of $B^H$, which is given by:

\[
\kappa_{\chi}(B^H) = \frac{\min \{Q^HQ^H\} + \xi}{\|Q^H\|^2 + K\xi}.
\]

Since for any $\xi \geq 0$ we have

\[
\frac{\min \{Q^HQ^H\}}{\|Q^H\|^2} \leq \frac{\min \{Q^HQ^H\} + \xi}{\|Q^H\|^2 + K\xi},
\]

it is possible to verify the slowest convergence occurs for $\xi = 0$. Thus, from [12, Proposition 2], it results in a theoretical upper bound in the number of iterations needed for the full convergence of the algorithm, i.e., $T_{\text{full convg}}$, which is given by:

\[
T_{\text{full convg}} \leq \frac{\|Q^H\|^2}{\min \{Q^HQ^H\}}.
\]

Hence, an order of $O\left(\frac{\|Q^H\|^2}{\min \{Q^HQ^H\}}\right)$ iterations is sufficient for the rKA attain full convergence, considering any regularization factor $\xi \geq 0$. 

\[
\text{Figure 3. Sum-Rate Capacity of the rKA, ZF and RCI precoders, considering EPA, } M \in [64, 128], K \in [32, 64], \tau = 0, T = 400 \text{ for } K = 32 \text{ and } T = 600 \text{ for } K = 64
\]
Also according to [12], when the channel vectors of the users have no spatial correlation, the full convergence upper bound can be approximated as:

\[ T_{\text{full convg}} \leq \frac{K}{(1 - \sqrt{\beta})^2}. \]  

(46)

1) rKA Convergence under Spatial Correlation, Channel Estimation and Increasing Number of Antennas: Fig. 4 suggests the rKA’s full convergence behavior w.r.t. the spatial antenna correlation \( c \) and channel estimation error \( \tau \), evaluated through (45) for a medium loading system in a massive MIMO configuration. The convergence impacts directly on the algorithm’s complexity. As expected, the number of iterations increases with the spatial correlation index but, interestingly, decreases with the channel estimation error. Moreover, as the number of BS antennas \( M \) and/or number of users \( K \) increases, the signal processing burden grows and, as a consequence, the full convergence of the rKA algorithm requires a higher number of iterations.

Fig. 5 presents the theoretical convergence results calculated both through the exact expression (45) and the approximation of (46). Fig. 5.a) shows that the theoretical \( T_{\text{full convg}} \) values present a linear dependence on the number of BS antennas. However, despite following the same pattern of growth of the precise results, the approximations implicate in a superior number of iterations needed for the full convergence of rKA. Furthermore, in Fig. 5.b) we see that the approximation of the theoretical results, evaluated by (46), indicates the need for a larger number of iterations than the exact calculations of (45) for channel correlation conditions in which \( c < 4 \). After that, the precise theoretical results of \( T_{\text{full convg}} \) exceed the approximations. It is also worth noting that the exact results increase exponentially w.r.t. the spatial antenna correlation.

Fig. 6 demonstrates that the theoretical UBs in the number of iterations becomes larger as the loading factor \( \beta \) increases. As in the previous figure, the approximated \( T_{\text{full convg}} \) values are shown to be higher than the more accurate ones, since an uncorrelated scenario was considered. Also, the difference among the curves seems to become more significant for higher values of \( \beta \).

Fig. 7 depicts the numerical system capacity results obtained for the rKA precoder vs the number of rKA iterations \( T \), for \( M = 64, K = 32 \), different spatial ULA correlation \( c \) and channel estimation conditions \( \tau \). Following Eq. (45), the upper bound in the number of iterations needed for the convergence of rKA for \( M = 64, K = 32 \) and an uncorrelated scenario is about \( T \approx 286 \) iterations, while for (46) it is of about 373 iterations. Analyzing the results from this figure we can see that the convergence seems to occur between 250 and 300 iterations. A similar number of iterations is needed for the convergence under low channel correlation conditions, e.g., \( c = 0.2 \), and slightly higher values of \( T \) are required for \( c = 0.5 \). Also, Fig. 7 indicates that the number of rKA iterations required for the full convergence of the rKA precoder, i.e., \( T_{\text{full convg}} \), is progressively reduced as the channel estimation quality parameter becomes larger, with each increase of 0.2 in the value of \( \tau \) corresponding to a reduction of about 50 iterations for the full convergence of the algorithm.
The trend is the same when the number of antennas $M$ increases but the system loading holds ($\beta = 0.5$). In the case of $M = 128$, applying (45) under different correlation and channel estimation conditions, the upper bound in the number of iterations needed for the rKA to achieve full convergence in an uncorrelated scenario is about 628 iterations. Meanwhile, this number is of about 746 iterations for the approximation of (46). The numerical results confirmed that the convergence occurs for 15 dB of SNR in the range of 550 to 600 iterations. Furthermore, it is straightforward to verify numerically (in same way found in Fig. 7) and analytically that the algorithm converges faster for higher values of $\tau$, i.e., for lower channel estimation quality conditions. Even though the rKA’s full convergence occurs under a higher number of iterations, increasing in 0.2 the value of $\tau$ results in a reduction of $T_{\text{full convg}}$ in approximately 150 iterations.

Fig. 8 shows the number of iterations for the full convergence of the rKA obtained both theoretically, using (45), and through observation of the numerical results for SNR = 15 dB, such as those presented in Fig. 7. The figure provides a comparison of both results obtained for $\beta = 0.5$, $c = 0.5$ and system dimensions $M = 64$ and $M = 128$.

As can be verified through the figure, the $T_{\text{full convg}}$ values of both results are quite similar for lower $\tau$, i.e., for superior channel estimation conditions. However, although both of them present a reduction in the number of iterations for higher values of $\tau$, this effect is much more evident in the numerical results. For the theoretical results, the reduction in the number of iterations for higher $\tau$ becomes more noticeable for $c \geq 0.5$.

2) rKA Convergence w.r.t. SNR: As can be seen from (43), the number of iterations needed for the full convergence of the rKA precoder is dependent of the regularization parameter $\xi$. Since in this work we consider this term inversely proportional to the SNR, we can conclude the value of SNR also impacts on the $T_{\text{full convg}}$.

Table III indicates that there is a dependence of $T_{\text{full convg}}$ w.r.t. SNR, where a higher number of iterations is required for higher SNR conditions. It can also be verified that the con-
vergence conditions are highly sensitive to an SNR alteration, since an increase in 5 dB of this value results in an increment of $T_{\text{full convg}}$ in approximately 150 iterations, for $M = 64$, and of $\approx 300$ iterations, for $M = 128$.

Examining the table below, it is possible to verify that the exact theoretical results are generally in accordance with the numerically obtained values for an SNR regime of 15 dB. However, these upper bounds are inferior to the numerical verifications for higher SNR regimes, such as 20 dB, for low correlation conditions, in which the approximated convergence results seem to represent a more adequate upper bound. On the other hand, the approximation results do not represent the convergence behavior of rKA for medium to high correlation scenarios, in which the evaluations from the exact expression appear to be more suitable to represent an upper bound for $T_{\text{full convg}}$.

Moreover, the numerical convergence of the rKA algorithm occurs for a similar number of iterations for the three correlation conditions analyzed, i.e., $c \in [0; 0.2; 0.5]$. However, $T_{\text{full convg}}$ appears to be slightly higher for $c = 0.5$, indicating an increase in the number of iterations for medium to high correlation conditions, i.e., for $c \geq 0.5$. This result is also in accordance with the behavior presented by the analytical results evaluated through (45).

\section*{D. Capacity under Equal Power Allocation}

The numerical results and bounds for the sum-rate capacity of the MIMO precoders are presented and compared under equal power assignment. The system analysis is performed as a data transmission from a BS with $M$ antennas to $K$ users (MT’s) equipped with a single antenna, characterizing the DL scenario of a massive MIMO single-cell.

First we have compared the sum-rate capacity attained with the classical ZF and RCI precoders, as well as the rKA precoder, under perfect CSI knowledge at the transmitter side for all cases. After that, we analyze the capacity degradation due to the noisy and possibly quantized version of the channel matrix, as described by Eq. (8), for different $\tau$'s.

1) Sum-Rate under Perfect CSI: Fig. 9 depicts the sum-rate capacity of ZF, RCI and rKA varying with the loading factor $\beta$ for $M = 128$ antennas and SNR $\gamma = 15$ dB. In such system operation conditions it can be seen that, despite the performance of RCI being always equal or better than that obtained by rKA or ZF precoding, the three precoding schemes result in a similar performance for $\beta \leq 0.3$. For higher values of the loading factor, however, we can see that the ZF presents progressively more degradation than the other two precoding schemes, what leads to a considerably inferior performance for $\beta$ close to the unity.

Regarding the behavior of rKA for $T_s$, the results show that, for $\beta \leq 0.2$, even $T = 200$ achieves the full convergence of the algorithm. As the value of $\beta$ increases to 0.3 and 0.4, respectively, the full convergence appears to occurs at about $T = 300$ and $T = 100$. Therefore, the increase in the cell load conditions leads to a need for a higher number of iterations in order to achieve the full convergence of rKA.

![Figure 9. Sum-Rate Capacity of the rKA, ZF and RCI precoders for EPA, $M = 128$, $\tau = 0$, $c = 0$, and $\beta \in [0, 1]$](image-url)

2) Sum-Rate under Imperfect CSI: In this subsection, we focus on the sum-rate capacity of the ZF, RCI and rKA precoders under different quality levels of CSI at the BS, i.e., assuming $\tau \in [0.0, 0.1, 0.2]$ in the channel model of eq. (8).

Fig. 10 depicts results for $M = 128$, $K = 64$ and correlation parameters $c = 0$ and $c = 0.5$. It can be seen that, under low correlation and for the considered number of iterations of rKA, both precoders perform almost identically in all the SNR range, for any $\tau$. Furthermore, RCI precoder presents a similar performance degradation than that experienced by the rKA, even under less accurate channel estimates ($\tau = 0.2$). When the level of channel correlation increases, e.g., $c = 0.5$, the sum-rate capacity for both precoders is reduced accordingly. The capacities achieved by RCI and rKA are similar to those obtained under low-correlation scenario. Also, the sum-rate capacity attained by deploying the RCI precoding degrades with the increase of $\tau$ in a similar way to the one seen with the rKA for the same channel estimation quality conditions.

On the other hand, even though it presents inferior sum-rate capacity results than the other schemes for lower SNR values, the ZF reaches the same sum-rate performance of RCI and rKA for higher SNR values. Moreover, under less accurate channel estimates (higher values of $\tau$), ZF and RCI precoders present a similar performance degradation than that experienced by the rKA.

Fig. 11 give us a general idea on how the sum-rate capacity results of the three precoding schemes is degraded w.r.t the increase of $\tau$ under low to high SNR regime. Since the condition of $M = 128$, $\beta = 0.5$ and SNR = 20 dB was also analyzed and Table III shows the numerical convergence for this specific scenario happens in between 850 and 900 rKA iterations, a number of $T = 900$ iterations was considered for the rKA while obtaining the results presented below.

In the different system dimensions and channel conditions considered, the capacity results are reduced for more spatially correlated channels. This degradation is not significant from
Table III
FULL CONVERGENCE (\(T_{\text{full convg}}\)) OF rKA PRECODING w.r.t. SNR REGIME UNDER \(\beta = 0.5\) AND \(\tau = 0\)

<table>
<thead>
<tr>
<th>Antennas (M)</th>
<th>Correlation (c)</th>
<th>SNR Regime (Simulation)</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>0</td>
<td>250 – 300 iter.</td>
<td>400 – 450 iter.</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>250 – 300 iter.</td>
<td>400 – 450 iter.</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>300 – 350 iter.</td>
<td>450 – 500 iter.</td>
</tr>
<tr>
<td>128</td>
<td>0</td>
<td>550 – 600 iter.</td>
<td>850 – 900 iter.</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>550 – 600 iter.</td>
<td>850 – 900 iter.</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>600 – 650 iter.</td>
<td>900 – 950 iter.</td>
</tr>
</tbody>
</table>

Figure 10. Sum-Rate Capacity of the rKA, ZF and RCI precoders for EPA, \(M = 128, K = 64\); \(\tau \in [0.0, 0.1, 0.2], c \in [0.0, 0.5]\), \(T = 400\) for \(K = 32\) and \(T = 600\) for \(K = 64\).

Figure 11. Sum-Rate degradation of the rKA, ZF and RCI for \(M = 128, K = 64, c = 0.5\) and \(\tau = 900\); equal power allocation.

zero to low channel correlation, i.e., from \(c = 0\) to \(c = 0.2\); the capacities for both conditions presented to be quite close to one another. The performance reduction is, however, much more noticeable in medium correlation conditions, i.e., \(c = 0.5\).

Another result verified in all the analyzed situations is an inferior performance of the ZF in comparison to RCI and rKA, for all values of \(\tau\) and SNR considered and specially for lower channel estimation quality and poor channel conditions. Also, the RCI and rKA precoders present a highly similar performance for all the analyzed situations.

In relation to the degradation of rKA and RCI for the introduction of medium correlation condition shown in the Figure, these two precoding schemes present, in a SNR of 0 dB, a degradation of about 9.1%, 11.1% and 12.5% for \(\tau = 0, \tau = 0.4\) and \(\tau = 0.8\), respectively. For a SNR = 10 dB regime, the reductions in the capacity results from the zero and low to medium correlation conditions are of about 11.1%, 17.1% and 19.1% for \(\tau = 0, \tau = 0.4\) and \(\tau = 0.8\), respectively. Likewise, for a SNR of 20 dB the degradation is around 14.3%, 15.91% and 20%, respectively, for \(\tau = 0, \tau = 0.4\) and \(\tau = 0.8\). Thus, in general, the capacity degradation of rKA and RCI precoders w.r.t an increase in the correlation appears to be higher for better channel conditions, as well as for higher values of \(\tau\).

With regard to the degradation of ZF, from zero/low to medium correlation conditions and a SNR of 0 dB, the capacity is reduced in 23.5%, 21.4% and 18.2% for \(\tau = 0, \tau = 0.4\) and \(\tau = 0.8\), respectively. For the same conditions as mentioned above and an SNR = 10 dB regime the degradations are over 15.4%, 14.3% and 14%. Besides, for SNR = 20 dB these reductions in the capacity are of approximately 14.3%, 20.4% and 14.3% for \(\tau = 0, \tau = 0.4\) and \(\tau = 0.8\), respectively. Therefore, the degradation in the performance of ZF precoding appears to be higher as the channel conditions become worse. With respect to the degradation in the channel estimation quality, the performance reduction of ZF seems to be similar or lower for an increment in the values of \(\tau\).

Thus, for worse channel conditions, the degradation provoked by an increase in the channel correlation seems to affect more significantly the ZF precoder. For better channel conditions, however, the rKA and RCI precoders have a more perceptible performance reduction occasioned by the more correlated conditions. Moreover, high channel estimation quality conditions associated with an increase in the correlation appear to interfere more in the performance of ZF, while the correspondent degradation of rKA and RCI seems to be more expressive for low channel estimation quality scenarios.
E. Precoders Complexity

Linear precoding techniques are computationally more efficient than its non-linear counterparts. However, most linear precoding schemes, such as the ZF and the RCI, present a number of arithmetic operations of $O(K^2 M)$. This leads to an intractable complexity for most of the linear precoding schemes in the massive MIMO regime [26].

Herein, we analyze the complexities of the RCI, ZF and rKA precoding schemes. These complexities are treated simply as the number of complex addition and multiplication operations needed for a given arithmetic operation [26]. Since the number of floating point operations (flops) needed to implement these complex operations varies according to the deployed hardware and complex number representation, we will not compare the number of flops of the implemented operations. Instead, we focus on the number of complex operations required for each linear precoder execute its function.

One way of obtaining the number of complex operations in order to calculate (20) consists of the following steps [26]:

1) compute the matrix-matrix multiplication $HH^H$;
2) add the diagonal matrix $\lambda$ to the result;
3) compute $H^H (HH^H + (\mathbf{I}_K))^{-1}$.

The number of complex operations needed to execute these steps is $K^2 (2M - 1) + K + \frac{4K^3}{3} + 2K^2 M$. Using the Choleski factorization to compute step 3) instead of completely inverting the matrix and, then, using matrix-matrix multiplication leads to the total cost of

$$C_t = K^2 (2M - 1) + K + \frac{4K^3}{3} + MK(2K - 1)$$

Furthermore, the RCI precoding matrix $\mathbf{G}_{\text{RCI}}$ is computed once per coherence time period. Once it has been obtained, the multiplication $\mathbf{G}_{\text{RCI}} \mathbf{x}$ considering $\mathbf{x} = \mathbf{I}_K$, which is computed every time the channel is used for DL transmission, requires extra $M(2K - 1)$ operations. Thus, the number of operations is in the order of $O(K^2) + O(K^2 M) + O(M K)$, which are extremely computationally demanding in large MIMO systems. Besides, the known inversion algorithms are complicated to implement in hardware.

The above complexity analysis can also be applied to the ZF precoding scheme. However, in the ZF case, $\xi = 0$; hence, step 2) is not necessary, what leads to a reduction of $K$ complex operations in the ZF if compared to RCI's complexity.

Considering the rKA precoder, as presented in the subsection V-C, a number of iterations in the order of $O\left(\frac{M||Q||^2}{\lambda_{\text{min}}(Q^H Q)}\right)$ is sufficient for the rKA computation in $M$ antennas. Indeed, since each step of KA requires $M$ multiplications, the computational complexity of KA for the described assumptions scales by $M$, where $O\left(\frac{M||Q||^2}{\lambda_{\text{min}}(Q^H Q)}\right)$ is the complexity per antenna. Besides, $O(M K)$ operations must be added for computing the sampling probability distribution and $O(M K)$ operations are needed for calculating $Q^H w$ in Algorithm 1, resulting in a computational complexity of $O\left(\frac{M||Q||^2}{\lambda_{\text{min}}(Q^H Q)}\right) + 2O(M K)$ for every channel use for data transmission. If the matrix $\mathbf{W}$ is calculated, then it is beneficial to keep the precoding matrix $\mathbf{G}_{\text{rKA}}$ in the factorized form $\mathbf{G}_{\text{rKA}} = a\mathbf{Q}^H \mathbf{W}$ and compute first the $\mathbf{W}$s multiplication of $\mathbf{x} = a\mathbf{Q}^H \mathbf{W}s$, which makes it possible to avoid an increase by a factor $O(K)$ in the complexity occasioned by the direct computation of $\mathbf{G}_{\text{rKA}}$. Therefore, the rKA presents an overall computational complexity much lower than the one associated with RCI and ZF precoders. The complexities of the analyzed schemes are included in the column “Complexity” of Table I.

VI. CONCLUSIONS

We have analyzed the sum-rate capacity achieved by three low-complexity linear precoders in a single-cell massive MIMO environment equipped with ULA antennas, considering the impact of correlation channels and imperfect channel state information. The ZF, RCI, and the iterative randomized Kaczmarz algorithm (rKA) precoding schemes were compared in terms of sum-rate capacity vs complexity. Numerical results demonstrated that, although the convergence of rKA presented to increase with the SNR, spatial channel correlation and cell load conditions, its performance-complexity tradeoff is much more favorable than that of RCI and ZF precoding when operating in realistic massive MIMO scenarios and conditions. This is due to rKA's relative low-complexity, which does not depend on the channel inversion. Besides, the rKA precoder demonstrated to be much more robust than ZF to the system loading increase (or interference), and equally efficient if compared to RCI under a simple equal power allocation policy. Moreover, imperfect channel estimates degrades similarly the performance of the three precoding techniques.

APPENDIX A

CAPACITY BOUNDS ACHIEVED WITH RKA PRECODING

Lower and upper bounds of the achievable ergodic rate of the rKA are presented in the following [12].

First, consider an arbitrary SLE $\mathbf{A}\mathbf{x} = \mathbf{b}$, and let $x^t$ be an estimation of rKA at iteration $t$ starting from the zero initialization $x^0 = 0$. Then, it is demonstrated in [12, Proposition 4] that $x^t = \mathcal{G}^t(\mathbf{A})\mathbf{b}$, where $\mathcal{G}^t(\mathbf{A})$ is a linear operator that depends on $\mathbf{A}$ and on the internal randomization of KA until iteration $t$, but does not depend on $\mathbf{b}$. The proof of this proposition is also presented in Appendix B.

The following lower and upper bounds on the achievable ergodic rate when the rKA is run for a specific number of iterations $t$ are derived assuming this Proposition.

In the DL, the signals received by the users are given by the vector $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ as described in Section II. From the proposition described above it results that the signal obtained by the rKA (Algorithm 1) at a specific iteration $t$ will be a random linear function of $\mathbf{s}$, such as:

$$\hat{\mathbf{z}} = \mathcal{G}^t(\mathbf{Q})\mathbf{s},$$

which directly leads to

$$\mathbf{w} = \mathcal{G}^t(\mathbf{Q})\mathbf{s}.$$  

According to (26) we can write

$$\mathbf{x} = \mathbf{Q}^H \mathcal{G}^t(\mathbf{Q})\mathbf{s} = \mathcal{G}^t(\mathbf{Q})\mathbf{s},$$

(49)
where the resultant $M \times K$ matrix $\mathcal{G}(Q)$ depends on the noisy channel state $Q$ and the internal randomization of $KA$, but does not depend on the signal $s$.

Since we also have

$$x = Gs = Q^HWs,$$

we obtain that $\mathcal{G}(Q) = G \in \mathbb{C}^{M \times K}$. Letting $\mathcal{G}(Q) = [g^1 \ldots g^K] \in \mathbb{C}^{M \times K}$, we can then interpret $g^k \in \mathbb{C}^{M \times 1}$ as the precoding vector corresponding to the $k$-th user at $t$-th iteration.

The time-varying\(^3\) interference matrix can be written as:

$$\mathbf{T}^0 = \mathbf{H}^\dagger(Q) = \mathbf{H}G \in \mathbb{C}^{K \times K},$$

where for simplicity we dropped the explicit dependence of $\mathbf{T}^0$ on the iteration $t$.

In the following, we use the upper and lower bounds on the capacity developed in [12], which in fact were developed in [25]. By treating the interference as noise and by coding across several coherence blocks, an upper bound on the ergodic capacity can be expressed as:

$$\overline{R}_k = \mathbb{E} \left[ \log_2 \left( 1 + \frac{|\mathbf{T}_{kk}^0|^2}{\sum_{k' \neq k} |\mathbf{T}_{k'k}^0|^2 + \sigma_k^2} \right) \right]$$

where the expectation is taken over all the randomness of the channel state $\mathbf{H}$ and the precoding matrix $\mathcal{G}(Q)$ including the randomness due to $KA$. This yields the upper bound $\overline{R} = \sum_{k=1}^K \overline{R}_k$ on the achievable sum rate.

For the lower bound on the ergodic rate, from [25], follows:

$$\underline{R}_k = \mathbb{E} \left[ \log_2 \left( 1 + \frac{|\mathbf{T}_{kk}^0|^2}{\sum_{k' \neq k} \mathbb{E}[|\mathbf{T}_{k'k}^0|^2] + \mathbb{E}[\sigma_k^2]} \right) \right]$$

which leads to the lower bound $\underline{R} = \sum_{k=1}^K \underline{R}_k$ on the ergodic rate.

\(\text{APPENDIX B}

\text{PROOF OF [12, PROPOSITION 4]}

We first assume that $x^t = \mathcal{G}^t(A)b$ is true for the $t$-th iteration. For $t = 0$, we have $x^0 = 0$ which is a zero function of $b$. Denoting $r(t) = 1, \ldots, m$ as the index of the row of $A$ selected at iteration $t$ and $a_r(t) = (A_r(t))^H$ we have:

$$x^{t+1} = x^t + \frac{b(t) - \langle a_r(t), x^t \rangle}{\|a_r(t)\|^2} a_r(t)$$

\((i)\) =

$$\begin{align*}
&= \frac{1}{\|a_r(t)\|^2} \left( a_r(t) - \langle a_r(t), e_r(t) \rangle \right) x^t + \frac{a_r(t) e_r(t)}{\|a_r(t)\|^2} b \\
&\overset{(ii)}{=} \frac{1}{\|a_r(t)\|^2} \left( a_r(t) - \langle a_r(t), e_r(t) \rangle \right) \mathcal{G}^t(A) + \frac{a_r(t) e_r(t)}{\|a_r(t)\|^2} b \\
&\overset{(iii)}{=} \mathcal{G}^{t+1}(A)b,
\end{align*}$$

where in $(i)$, $e_r(t) \in \mathbb{C}^{m \times 1}$ is the canonical vector with “1” as the index $r(t)$ and “0” leading to $b_r(t) = e_r(t) b$, and in $(ii)$ the induction hypothesis that $x^t = \mathcal{G}^t(A)b$ is used. This implies that $x^{t+1} = \mathcal{G}^{t+1}(A)b$, where the linear operator $\mathcal{G}^{t+1}(A)$ depends on $A$ and on the internal randomization of $KA$ until iteration $t + 1$ but does not depend on $b$.

\(\text{APPENDIX C}

\text{ZF AND RCI SUM-RATE UNDER LARGE SYSTEM REGIME}

Let’s consider the SINR formulation for the RCI precoder in Eq. (25), with a finite size system and path-loss coefficients to each user. Such consideration leads to each user having specific path-loss values. We can then rewrite (25) as:

$$\text{SINR}^\text{RCI}_k = \frac{\alpha^2 a_k^2 \mathbf{S}_k}{\alpha^2 a_k^2 I_k + \sigma_n^2} = \frac{\alpha^2 a_k^2 |A_k|^2}{\alpha^2 a_k^2 I_k + \sigma_n^2}$$

where $A_k = h_k (H_k^H H_k + \xi I_M)^{-1} H_k^H$ and

$$I_k = \sum_{j \neq k} |h_k (H_k^H H_k + \xi I_M)^{-1} h_j H_j|^2.$$

where $\alpha^2$, $A_k$ and $I_k$ are random quantities which depend on $H[11]$.

Let us assume $H = \sqrt{\rho} \hat{H}$ and $h_k = \sqrt{\rho} \hat{h}_k$, where, as in (4), $\hat{h}_k \sim \mathcal{CN}(0_{1 \times M}, I_M)$ is the uncorrelated small-scale channel fading vector for the user $k$. Besides, $A_k$ and $I_k$ can then be rewritten respectively as:

$$A_k = \hat{h}_k (H_k^H H_k + \rho \Phi^{-1})^{-1} H_k^H$$

and

$$I_k = \sum_{j \neq k} |h_k (H_k^H H_k + \rho \Phi^{-1})^{-1} h_j H_j|^2.$$

In the large system regime, $A_k$ converges almost surely to the deterministic quantity $\nu$ [24], which can be found through the expression:

$$\nu = \rho(1 + \nu)^2 E_{12} + \beta E_{22} + \beta \nu E_{22},$$

where $\rho = \xi/M$ is the normalized regularization parameter and $E_{ij}$ is given by [24]

$$E_{ij} = \mathbb{E} \left[ \frac{T^i}{(\rho(1 + \nu) + \beta T)^j} \right],$$

where the expectation is taken over the random variable $T$ whose distribution function is the limiting eigenvalue distribution of the correlation matrix $\Phi$.

The interference term $I_k$ converges almost surely (a.s.) to:

$$I_k \overset{a.s.}{\to} \frac{\beta E_{22}}{(1 - \beta E_{22}) (1 + \nu)^2}.$$ (60)

Besides, the term $\alpha^2$ can be shown to converge a.s. to:

$$\alpha^2 \overset{a.s.}{\to} \frac{P_k (1 - \beta E_{22})}{\beta E_{12}}.$$ (61)

Combining the large system limit results for $A_k$, $I_k$ and $\alpha^2$, the SINR of user $k$ converges almost surely to the limiting deterministic SINR:

$$\text{SINR}^{\text{RCI}}_k = \frac{\gamma_k \nu^2 (1 - \beta E_{22})}{\gamma_k E_{22} + (1 + \nu)^2 E_{12}}$$ (62)
where $\gamma_k = \frac{P \sigma_n^2}{\sigma^2}$ is the effective SNR.

According to the exponential correlation model presented in Section II, eq. (6), we calculate $E_{12}$ and $E_{22}$ as in [24, eq. (4.20) and (4.21)]. Based on these considerations, $\nu$ can be obtained through:

$$\nu = \sqrt{\left(\frac{1+c+\beta}{1+c+\beta}\right) + \frac{\beta(1+c^2)}{1+c^2}}.$$  \hspace{1cm} (63)

Also, the limiting SINR on (62) can be rewritten as

$$\text{SINR}_{\text{linear}} = \frac{\gamma_k \Theta_1 + \sigma_n^2 (1+\nu)^2 \Theta_2}{\gamma_k \Theta_1 + (1+\nu)^2 \Theta_2},$$ \hspace{1cm} (64)

where $\Theta_1 = \rho(1+\nu)(1+c^2) + \beta(1-c^2)$ and $\Theta_2 = \rho(1+\nu)(1-c^2) + \beta(1+c^2)$.

For the ZF precoder, the limiting SNR under the considered correlation model can be expressed as [24]:

$$\text{SINR}_{\text{ZF}} \approx \frac{\gamma_k (1-\beta^2)}{\gamma_k (1-\beta^2) + \mu^2 \sigma^2} < 1 \hspace{1cm} (65)$$

where $\mu = \frac{1+c^2}{\sqrt{c}}$.

REFERENCES


Bibliography


