Resource Allocation and Pairing Techniques in Multiuser Massive MIMO-NOMA

Eric Pierre Simon, Joumana Farah, Pierre Laly

Resource Allocation and Pairing Techniques in Multi-User Massive MIMO-NOMA

Eric Pierre Simon[1], Joumana Farah[2], Pierre Laly[1]

***Abstract*—In massive multiple-input multiple-output (MIMO) systems, user clustering techniques are crucial for addressing inter-beam interference. These techniques become even more critical when incorporating non-orthogonal multiple access (NOMA) to enhance spectral efficiency and user fairness. Prior research in this area has primarily focused on channel correlation for user pairing. This paper presents a novel approach for user pairing and subband allocation in a crowded downlink system that utilizes a criterion based on minimizing the condition number (CN) of channel matrices. A benchmark resource allocation technique, which is based on a rate maximization criterion and has a higher complexity but quasi-optimal performance, is also introduced. To evaluate the performance of the proposed technique in a realistic setting, experimental massive MIMO channel measurements in a dense user deployment scenario are used. Results indicate that the CN criterion leads to significantly higher throughput and fairness levels compared to the channel correlation criterion and achieves performance that is close to that obtained with rate maximization. The study is further extended to the case of multi-antenna reception, where interference cancellation techniques are proposed to increase the throughput of users with low channel gains. The proposed methods outperform a well-referenced technique in the literature.**

***Index Terms*—Massive Multiple-Input Multiple-Output, Non-Orthogonal Multiple Access, User Pairing, Subband Allocation, Channel Measurements.**

I. Introduction

The increasing demand for data in communication systems, driven by the growing popularity of video streaming, teleworking, smart cities, e-health, and connected vehicles, poses new challenges and constraints for researchers and engineers, particularly in terms of limited spectrum availability and high levels of interference. Massive multiple-input multiple-output (MIMO) is a promising technique in 5G and beyond systems for improving the performance of previous networks, as demonstrated in studies such as [1], [2]. However, its implementation is limited by the number of users that can be efficiently served on a common time-frequency resource through precoding. Research has shown that the recommended ratio of antennas to beams is commonly in the range of 2 to 8 [3], [4].

A. Background and related work

To address the aforementioned limitations of MIMO systems, non-orthogonal multiple access (NOMA) techniques can be incorporated into massive MIMO systems, as reported in previous studies such as [5]–[8]. NOMA is achieved by multiplexing two or more signals in the power domain on the same beam, with the use of successive interference cancellation (SIC) at some of the receivers to separate the signals.

Previous works of the literature have shown that power-domain NOMA can significantly outstrip the system performance of its orthogonal multiple access (OMA) counterparts [5], [9], [10]. However, it is important to note that this improvement can only be realized when coupled with efficient resource allocation and beamforming techniques, as reported in various studies such as [11]–[17]. For instance, in [11], system capacity is maximized by optimizing the beamforming and power allocation. In [12], a signal alignment method for MIMO-NOMA is proposed, which can be applied when the number of receive antennas exceeds half that of the transmitter. This framework is further revisited in [13], where an interference suppression method is proposed in such a way as to break the constraint on the number of antennas and allow its application in a massive MIMO context. In [14], an analysis of the best user multiplexing scheme is conducted based on user locations, with a preference for maximum ratio transmission (MRT) to limit complexity. In [15], a simple clustering solution is proposed where highly channel-correlated users with maximum channel gain difference are selected to be assigned to a single subband. A generalization of the above approach is proposed in [16] for multi-subband allocation with a variable number of clusters per subband and a moderate number of served users.

Recently, in [17], a greedy assignment technique of users to subbands and beams is introduced in the context of a crowded system. By gradually relaxing the correlation constraints, an important gain is achieved with NOMA towards OMA and state-of-the-art NOMA techniques.

B. Contributions and paper organization

In this paper, we present several resource allocation techniques that aim to effectively serve a large number of users over a limited number of subbands. One of the main contributions of this work is the use of the condition number (CN) as an efficient criterion for subband loading. We demonstrate that this criterion, when used effectively, can greatly overcome the correlation criterion that is widely used in most previous works in the field. Additionally, we show that the use of CN can yield system performance that is very close to that of the

[1]E.P. Simon and P. Laly are with Univ. Lille, CNRS, UMR 8520-IEMN laboratory, F-59000 Lille, France. e-mail: eric.simon@univ-lille.fr.

[2]J. Farah is with Univ Rennes, INSA Rennes, CNRS, IETR-UMR 6164, F-35000 Rennes, France.

total throughput criterion while having much lower complexity. This complexity is further reduced by incorporating an iterative matrix inversion technique. Our proposed resource allocation methods are applied to both OMA and NOMA cases, and we show how NOMA can further enhance system performance compared to OMA. We also adapt the allocation methods to the multi-antenna receiver case and introduce two new receive combining methods that significantly outperform the well-referenced method presented in [13]. Furthermore, the performance of the proposed methods is studied using real-world channel measurements rather than a simple simulated environment to account for the complexity of a realistic transmission environment.

In summary, this paper makes the following key contributions:

- the introduction of two algorithms for joint subband (frequency) and beam (spatial) resource allocation in highly populated massive MIMO systems,
- the first algorithm employs the CN as a metric for effectively assigning users to resources, ensuring a low complexity cost,
- the second algorithm utilizes the total rate as a metric, providing a benchmark for performance evaluation,
- the incorporation of NOMA through the development of novel pairing techniques,
- the extension of the proposed algorithms to the case of multi-antenna receivers by designing combining techniques that efficiently manage interference,
- the evaluation of the algorithms performance through practical measured channels.

The organization of this article is as follows: Section II presents the system model used in the study. The proposed resource allocation techniques, including the complexity reduction method, are discussed in detail in Section III. In Section IV, the channel measurements conducted to validate the study are described. Results and analysis of the proposed techniques are presented in Section V. Finally, the paper is concluded in Section VI.

II. System Model

The proposed system is a downlink Massive MIMO-NOMA system designed to serve a large number of K users randomly distributed within a cell. The serving base station (BS) is equipped with M antennas and transmits information to the users using a bandwidth equally divided into S subbands. Each subband comprises an integer number of OFDM (orthogonal frequency division multiplexing) subcarriers. A zero-forcing (ZF) precoding scheme is applied by the BS separately on each subband s, $s = 1, \ldots, S$, using B_s beams. In the NOMA scheme, two or more users are served on the same beam. The maximum number of users per beam is set to 2, as was done in previous studies such as [10], [12], [13], [18], [19]. Indeed, a greater number of accommodated users per cluster does not necessarily translate into higher throughput performance because of the generated inter-user (intra-beam and inter-beam) interference at the level of users that can not perform SIC. At the same time, this enlargement incurs a significant increase in complexity at the BS level as well as at the user equipment levels, with a higher risk of error propagation at the SIC receivers. Let $\mathbf{W}_s$ be the $M \times B_s$ precoding matrix of subband s, $\mathbf{w}_{n,s}$ the n^{th} column of $\mathbf{W}_s$, and $P_{n,s}$ the transmit power on the n^{th} beam of subband s. The total power transmitted by the BS is therefore $P_{\text{tot}} = \sum_{s=1}^{S} \sum_{n=1}^{B_s} P_{n,s}$. The BS transmits the combined signal on subband s:

$$\mathbf{x}_s = \sum_{n=1}^{B_s} \mathbf{w}_{n,s} x_{n,s}. \tag{1}$$

Each beam can either accommodate one or two users. For a two-user beam, the signal $x_{n,s}$ with power $E[|x_{n,s}|^2] = P_{n,s}$ is written as $x_{n,s} = \sqrt{\alpha_{n,s,1}} a_{n,s,1} + \sqrt{\alpha_{n,s,2}} a_{n,s,2}$, where $a_{n,s,1}$ and $a_{n,s,2}$ are the transmitted signals of the two paired users on beam n of subband s, $\forall n, s, i$. $\alpha_{n,s,1}$ and $\alpha_{n,s,2}$ are power coefficients verifying $\alpha_{n,s,1} + \alpha_{n,s,2} = 1, \forall n, s$. Considering first the case of single-antenna users, the received signal by the i^{th} receiver, $i = 1, 2$, is given by:

$$y_{n,s,i} = \mathbf{h}_{n,s,i} \mathbf{x}_s + q_{n,s,i}, \tag{2}$$

where $\mathbf{h}_{n,s,i}$ is the $1 \times M$ channel vector of user i on beam n of subband s. $q_{n,s,i}$ is an additive white complex Gaussian noise with zero mean and variance σ^2. Users in pairs are sorted by their channel gains such that $\|\mathbf{h}_{n,s,1}\| > \|\mathbf{h}_{n,s,2}\|$, and the user indexed by $i = 1$ (resp. $i = 2$) is denoted as the strong (resp. weak) user in the pair (cluster). In two-user beams, ZF precoding is done based on the channel gains of the strong users, as was done in [15] and [16]. Let $\mathbf{H}_s$ be the $B_s \times M$ matrix constituted by the channel vectors of the strong users within clusters and those of unique users in single-user beams. The ZF precoding matrix is calculated by:

$$\tilde{\mathbf{W}}_s = \mathbf{H}_s^H (\mathbf{H}_s \mathbf{H}_s^H)^{-1}, \tag{3}$$

which then undergoes a normalization of each beam precoding vector as follows:

$$\mathbf{w}_{n,s} = \frac{\tilde{\mathbf{w}}_{n,s}}{\|\tilde{\mathbf{w}}_{n,s}\|}, \tag{4}$$

where $\tilde{\mathbf{w}}_{n,s}$ is the n^{th} column of $\tilde{\mathbf{W}}_s$.

The equivalent channel gain of the i^{th} user on the two-user beam n of subband s is $g_{n,s,i,m} = \mathbf{h}_{n,s,i} \mathbf{w}_{m,s}$. When $m \neq n$, $|g_{n,s,i,m}|^2 P_{m,s}$ quantifies the amount of interference caused by beam m on beam n. Besides, $g_{n,s,i,n}$ is the useful channel gain of the i^{th} user on the n^{th} beam. For a beam with a single user of channel $\mathbf{h}_{n,s}$, the equivalent channel gain is $g_{n,s,m} = \mathbf{h}_{n,s} \mathbf{w}_{m,s}$. The strong user in a NOMA beam n can suppress intra-cluster interference (ICI) by SIC and subtract the signal of the weak user from the received signal $y_{n,s,i}$ before decoding its own signal. As for the weak user, it does not perform SIC [5], [10] and directly proceeds to the decoding of its signal while suffering from ICI and also from the inter-beam interference (IBI) incurred by signals from other beams. In this work, a perfect SIC is assumed at the level of strong users. The rates achieved by the strong and

weak users of the two-user beam n are, respectively:

$$R_{n,s,1} = \log_2\left(1 + \frac{|g_{n,s,1,n}|^2 P_{n,s}\alpha_{n,s,1}}{\sum_{m=1,m\neq n}^{B_s}|g_{n,s,1,m}|^2 P_{m,s} + \sigma^2}\right), \quad (5)$$

$$R_{n,s,2} = \log_2\Bigg(1+ \frac{|g_{n,s,2,n}|^2 P_{n,s}\alpha_{n,s,2}}{|g_{n,s,2,n}|^2 P_{n,s}\alpha_{n,s,1} + \sum_{m=1,m\neq n}^{B_s}|g_{n,s,2,m}|^2 P_{m,s} + \sigma^2}\Bigg). \quad (6)$$

On a single-user beam, the rate of the unique user is:

$$R_{n,s} = \log_2\left(1 + \frac{|g_{n,s,n}|^2 P_{n,s}}{\sum_{m=1,m\neq n}^{B_s}|g_{n,s,m}|^2 P_{m,s} + \sigma^2}\right). \quad (7)$$

Note that these general expressions are provided to incorporate any precoding scheme. In the case of ZF precoding, the IBI terms in Eqs. (5) and (7) are suppressed. However, second users still suffer from IBI since ZF precoding is designed based on first users only.

As for the case of multi-antenna receivers, the system model is modified to take into account the fact that $\mathbf{h}_{n,s,i}$ is now an $N \times M$ channel matrix, N being the number of receive antennas. An $N \times 1$ combining vector $\mathbf{z}_{n,s,i}$ is applied at each receiver: $\mathbf{z}_{n,s,i}^H \mathbf{y}_{n,s,i}$, $i = 1, 2$, where $\mathbf{y}_{n,s,i}$ is the $N \times 1$ received signal vector at the i^{th} user of the n^{th} beam. For first users, an equal gain combining (EGC) is assumed: $\mathbf{z}_{n,s,1} = \frac{1}{\sqrt{N}}[1\,1\,\cdots\,1]^T$, as was done in the well-referenced work from the literature [13]. Therefore, at the base station level, ZF precoding is still performed based on Eqs. (3) and (4), by using the averaged channel gains of the first users. While inter-beam interference is canceled for first users using ZF precoding, a different scheme needs to be applied in order to remove IBI at second users by exploiting the receiver spatial diversity through a proper combining of the received signals.

The authors of [13] propose to cancel IBI by first defining the following combining $N \times 1$ vector of the second user on the n^{th} beam on subband s:

$$\mathbf{z}_{n,s,2} = \mathbf{h}_{n,s,2}\mathbf{v}_{n,s,2}, \quad (8)$$

where $\mathbf{v}_{n,s,2}$ is an $M \times 1$ vector to be determined. However, we will show that the reference method does not efficiently resolve the IBI cancelation problem at second users. Indeed, based on the definition in Eq. (8), we can express the following processed received signal at a second user:

$$\mathbf{z}_{n,s,2}^H \mathbf{y}_{n,s,2} = \mathbf{v}_{n,s,2}^H \mathbf{h}_{n,s,2}^H \mathbf{h}_{n,s,2}(\mathbf{w}_{1,s}x_{1,s} + \cdots + \mathbf{w}_{n,s}x_{n,s} + \cdots + \mathbf{w}_{B_s,s}x_{B_s,s}) + \mathbf{v}_{n,s,2}^H \mathbf{h}_{n,s,2}^H \mathbf{q}_{n,s,2}, \quad (9)$$

where $\mathbf{q}_{n,s,2}$ is the $N \times 1$ noise vector at the second user of the n^{th} beam.

In our work, we define the following $M \times 1$ effective channel vector at second users:

$$\mathbf{g}_{n,s,2,k} = \mathbf{h}_{n,s,2}^H \mathbf{h}_{n,s,2}\mathbf{w}_{k,s}. \quad (10)$$

By plugging the notation in Eq. (10) into Eq. (9), the latter can be re-written as:

$$\mathbf{z}_{n,s,2}^H \mathbf{y}_{n,s,2} = \mathbf{v}_{n,s,2}^H(\mathbf{g}_{n,s,2,1}x_{1,s} + \mathbf{g}_{n,s,2,2}x_{2,s} + \cdots + \mathbf{g}_{n,s,2,n}x_{n,s} + \cdots + \mathbf{g}_{n,s,2,B_s}x_{B_s,s}) + \mathbf{v}_{n,s,2}^H \mathbf{h}_{n,s,2}^H \mathbf{q}_{n,s,2}. \quad (11)$$

Consequently, to cancel IBI at second users, the vector $\mathbf{v}_{n,s,2}$ should be found in such a way to guarantee the following equality:

$$\mathbf{v}_{n,s,2}^H[\mathbf{g}_{n,s,2,1}\,\mathbf{g}_{n,s,2,2}\,\cdots\,\mathbf{g}_{n,s,2,n}\,\cdots\,\mathbf{g}_{n,s,2,B_s}] = \mathbf{i}_n, \quad (12)$$

where $\mathbf{i}_n$ is a $1 \times B_s$ vector with zero values except at the n^{th} position where the value is 1.

The rate expressions (5), (6) and (7) become respectively (13), (14) and (15):

$$R_{n,s} = \log_2\Bigg(1+ \frac{|\mathbf{z}_{n,s}^H \mathbf{h}_{n,s}\mathbf{w}_{n,s}|^2 P_{n,s}}{\sum_{m=1,m\neq n}^{B_s}|\mathbf{z}_{n,s}^H \mathbf{h}_{n,s}\mathbf{w}_{m,s}|^2 P_{m,s} + \sigma^2\|\mathbf{z}_{n,s}\|^2}\Bigg), \quad (15)$$

where $\mathbf{z}_{n,s} = \mathbf{z}_{n,s,1}$.

In [13], a different combining approach is applied based on the following equivalent channel vector:

$$\mathbf{g}_{n,s,2} = \mathbf{h}_{n,s,2}^H \mathbf{h}_{n,s,2}\mathbf{w}_{n,s}. \quad (16)$$

By inspecting Eq. (14), it appears that the IBI is not canceled based on the definition in Eq. (16). This observation will be further investigated in the simulation results, where both combining methods will be tested and compared to another one that applies a simple EGC at second users.

In addition to our interference cancelation combining method, we also provide an MMSE approach:

$$\mathbf{z}_{n,s,2} = \left(\sum_{\substack{m=1 \\ m\neq n}}^{B_s} \mathbf{h}_{n,s,2}\mathbf{w}_{m,s}\mathbf{w}_{m,s}^H \mathbf{h}_{n,s,2}^H P_{m,s} + \sigma^2\mathbf{I}_N\right)^{-1} \mathbf{h}_{n,s,2}\mathbf{w}_{n,s}, \quad (17)$$

where $\mathbf{I}_N$ is the identity matrix of size $N \times N$.

III. Proposed resource allocation techniques

In this section, we first present two resource allocation techniques for the OMA scenario while considering single-antenna receivers. These techniques are then extended to the NOMA case. Subsequently, an adaptation of the proposed methods to multi-antenna receivers is carried out. The first proposed technique utilizes a decision metric based on the CN, while the second is based on the maximization of the total rate (TR). Although the second approach is quasi-optimal, it is more complex and mainly serves as a benchmark for the performance evaluation of the algorithms.

Our approach based on the CN is inspired by the observations in [20] where it was noted that precoding techniques that rely on matrix inversions like MMSE or ZF tend to achieve poor sum-rate performance when the CN of the channel gain matrix is large. This is due to the fact that a matrix channel realization with a high CN corresponds to the users' channel

$$R_{n,s,1} = \log_2\left(1 + \frac{|\mathbf{z}_{n,s,1}^H \mathbf{h}_{n,s,1}\mathbf{w}_{n,s}|^2 P_{n,s}\alpha_{n,s,1}}{\sum_{m=1,m\neq n}^{B_s}|\mathbf{z}_{n,s,1}^H \mathbf{h}_{n,s,1}\mathbf{w}_{m,s}|^2 P_{m,s} + \sigma^2\|\mathbf{z}_{n,s,1}\|^2}\right), \tag{13}$$

$$R_{n,s,2} = \log_2\left(1 + \frac{|\mathbf{z}_{n,s,2}^H \mathbf{h}_{n,s,2}\mathbf{w}_{n,s}|^2 P_{n,s}\alpha_{n,s,2}}{|\mathbf{z}_{n,s,2}^H \mathbf{h}_{n,s,2}\mathbf{w}_{n,s}|^2 P_{n,s}\alpha_{n,s,1} + \sum_{m=1,m\neq n}^{B_s}|\mathbf{z}_{n,s,2}^H \mathbf{h}_{n,s,2}\mathbf{w}_{m,s}|^2 P_{m,s} + \sigma^2\|\mathbf{z}_{n,s,2}\|^2}\right), \tag{14}$$

gain vectors being nearly linearly dependent. Furthermore, the CN metric was previously employed in [4] to determine the optimal ratio between the number of users and transmit antennas, thereby maximizing the spatial user separability. In a related study [17], it was demonstrated that the CN serves as a reliable indicator of system-level performance, particularly when the number of users per subband closely approaches the number of transmit antennas. Drawing from these previous findings, we have proposed our first algorithm for the OMA context, where we leverage the CN as a metric to allocate users to specific subbands.

A. OMA Algorithms

In the current part, we denote by $\mathbf{h}_{k,s}, k = 1, ..., K, s = 1, ..., S$, the $1 \times M$ channel vector of user k measured on the s^{th} subband. Indeed, at this point, the assignment of users to beams has not been performed yet; therefore, the beam index n is now discarded in the notation and replaced by the user index k. Furthermore, Matlab notation is used for matrices and vector indices within the algorithms' pseudo-codes.

The first technique, denoted by OMA-CN and presented in the pseudo-code of Algorithm 1, starts by sorting the priorities of all active users based on the channel gain difference between their best subband and their next best subband (steps 5 to 10): the first user k to be assigned is the one for which the second-best channel gain is much worse than the best one. This prioritization scheme aims at ensuring that users experiencing a poor channel condition get a higher chance of obtaining their best available subband. Consider, for example, the case of the transfer function of receiver 139 in Fig. 3, which shows a few measured channel transfer functions (as will be detailed in Section IV). We notice that this user experiences an overall difference of around 10 dB between its best and worst subbands. Generally, an important difference in channel gains reflects a significant difference in the signal-to-interference-and-noise ratio (SINR) and, subsequently, in the user's achieved rate. Therefore, this prioritization guarantees that a user gets its best subband in case it is the only available one with a good channel gain for this user, as was also noted in [9]. This is all the more important as the number of available subbands decreases throughout the iterative allocation process, where subbands are gradually loaded. However, since at the beginning of the algorithm, all tested subbands yield the same initial CN for a given user, the first S users are assigned to different subbands, which are chosen based on the maximum channel norm (steps 11 to 15). Afterward, the set of available subbands, i.e., the subbands with a current number of assigned users smaller than M, is explored for each user k. The latter is added to the subband for which its addition yields the least CN of the matrix $\mathbf{H}_s^{\text{test}}(\mathbf{H}_s^{\text{test}})^H$ (steps 16 to 28) whose inverse is used in the calculation of the precoding matrix. $\mathbf{H}_s^{\text{test}}$ is the matrix obtained by the vertical concatenation of the current channel matrix $\mathbf{H}_s$ with the channel vector of user k on the tested subband s.

The second proposed technique, named OMA-TR, differs from OMA-CN by the decision criterion, which is now based on the total achievable rate: user k is assigned to the subband for which this assignment yields the highest system throughput. This second approach will be used as a benchmark for performance comparison in Sections III-F and V.

B. Iterative method

Given that the number of matrix inversions required for beamforming and CN calculation increases significantly with the number of users, we propose reducing the computational burden by using an iterative inversion technique. This technique can be implemented whenever a new channel vector is added to the channel matrix to determine the new precoding matrix or CN. For example, refer to steps 19 and 20 in Algorithm 1. The CN calculation is performed using: $\text{cond}(\mathbf{X}) = \|\mathbf{X}\|\|\mathbf{X}^{-1}\|$. At a certain stage of the allocation process, $\mathbf{H}_s^{\text{test}}$ is of dimensions $i \times M$ and is denoted by $\mathbf{H}_s^{\text{test (i)}}$, where i is the actual number of beams on subband s. $\mathbf{H}_s^{\text{test (i)}}$ is then vertically appended by the channel vector $\mathbf{h}_{k,s}$ of size $1\times M$ to yield a new matrix $\mathbf{H}_s^{\text{test (i+1)}}$ of dimensions $(i+1) \times M$. Let $\mathbf{X}_i = \mathbf{H}_s^{\text{test (i)}}(\mathbf{H}_s^{\text{test (i)}})^H$ and $\mathbf{X}_{i+1} = \mathbf{H}_s^{\text{test (i+1)}}(\mathbf{H}_s^{\text{test (i+1)}})^H$ the square matrices of dimensions $i \times i$ and $(i+1) \times (i+1)$ respectively. We can express $\mathbf{X}_{i+1}$ in terms of $\mathbf{X}_i$ by:

$$\mathbf{X}_{i+1} = \begin{bmatrix} \mathbf{X}_i & \mathbf{H}_s^{\text{test (i)}}(\mathbf{h}_{k,s})^H \\ \mathbf{h}_{k,s}(\mathbf{H}_s^{\text{test (i)}})^H & \mathbf{h}_{k,s}(\mathbf{h}_{k,s})^H \end{bmatrix}. \tag{18}$$

Based on the Bordering Method [21], the inverse $\mathbf{X}_{i+1}^{-1}$ of $\mathbf{X}_{i+1}$ is related to $\mathbf{X}_i^{-1}$ by:

$$\mathbf{X}_{i+1}^{-1} = \begin{bmatrix} \mathbf{X}_i^{-1} + \frac{\mathbf{X}_i^{-1}\delta_i\delta_i^H\mathbf{X}_i^{-1}}{v_i} & -\frac{\mathbf{X}_i^{-1}\delta_i}{v_i} \\ -\frac{\delta_i^H\mathbf{X}_i^{-1}}{v_i} & \frac{1}{v_i} \end{bmatrix}, \tag{19}$$

with $\delta_i = \mathbf{H}_s^{\text{test (i)}}(\mathbf{h}_{k,s})^H$ and $v_i = \mathbf{h}_{k,s}(\mathbf{h}_{k,s})^H - \delta_i^H\mathbf{X}_i^{-1}\delta_i$.

In the methods OMA-CN-Iter and OMA-TR-Iter, Eq. (19) is used in step 20 of Algorithm 1 and in step 10 of Algorithm 2 to recursively estimate the CN or the precoding matrix after appending a new channel vector to the previous matrix.

C. NOMA Algorithms

It is well-known that limiting the number of beams per subband leads to better conditioning of the channel matrix [16], [17], [22] in both beamforming and CN calculation. This can be achieved by adopting NOMA, where a portion of the subband beams supports pairs of users instead of single users, resulting in improved performance through pairing. To this end, Algorithm 3 (NOMA-CN or NOMA-TR) extends our OMA algorithms to the NOMA scenario. The algorithm is based on the outcome of the previous OMA phase (either OMA-CN or OMA-TR). A vector $\mathbf{u}$ is then constructed (step 3) using the last $K/2$ users allocated on each subband in the OMA phase. The other allocated users on each subband s are kept on the beams assigned to them by the OMA phase. Then, each user in $\mathbf{u}$ is either paired on an appropriate beam and subband (steps 18 to 43, and 52 to 57) or reassigned as a first user to its original subband (steps 48 to 50), depending on the allocation scheme that yields the maximum throughput. Note that, in the NOMA phase, only the total rate criterion was used, being much more relevant than the CN at this stage since the size of the channel matrix of first users varies very slightly throughout the pairing process.

In all the proposed algorithms, an equal inter-beam power allocation of the total power budget P_{tot} was adopted. More elaborate power allocation techniques as the ones proposed in [10], [16], [19], [23] could be included within the proposed techniques. However, this is beyond the scope of this paper which focuses on channel allocation and user pairing.

Algorithm 2: OMA-TR: OMA with Total Rate Criterion

Input : M, K, S, $\mathbf{h}_{k,s}, k = 1, ..., K, s = 1, ..., S$, P_{tot}
Output: Assignment matrix $\mathbf{A}_{\text{OMA-TR}}$ of size $S \times M$

1 $\mathbf{A}_{\text{OMA-TR}} = \mathbf{0}$;
2 $n_s = 0$, $s = 1, \ldots, S$; // Variable containing the number of beams on subband s
3 $\mathbf{H}_s = []$, $s = 1, \ldots, S$; // Channel matrix of subband s
4 $R_s = 0$, $s = 1, \ldots, S$; // Total rate on subband s
5 Obtain the vector $\mathbf{k_d}$ of user indices sorted by priority as in steps 5 to 10 in Algorithm 1;
6 **for** $k = \mathbf{k_d}(1) : \mathbf{k_d}(\textit{length}\ (\mathbf{k_d}))$ **do**
7 **for** $s = 1 : S$ **do**
8 **if** $n_s < M$ **then**
9 $\mathbf{H}_s^{\text{test}} = [\mathbf{H}_s; \mathbf{h}_{s,k}]$;
10 Calculate the precoding matrix $\mathbf{W}_s^{\text{test}}$ for matrix $\mathbf{H}_s^{\text{test}}$ using Eqs. (3) and (4);
11 $R_s^{\text{test}} = \sum_{n=1}^{n_s+1} R_{n,s}$, $R_{n,s}$ given by Eq. (7);
12 $R_s^{\text{tot}} = R_s^{\text{test}} + \sum_{s' \neq s} R_s$;
13 **end if**
14 **end for**
15 $s^* = \text{argmax}_{s, n_s < M} R_s^{\text{tot}}$;
16 $\mathbf{H}_{s^*} = \mathbf{H}_{s^*}^{\text{test}}$; $R_{s^*} = R_{s^*}^{\text{test}}$;
17 $n_{s^*} = n_{s^*} + 1$; $\mathbf{A}_{\text{OMA-TR}}(s^*, n_{s^*}) = k$;
18 **end for**

D. Comparison with the correlation criterion

The proposed methods will be compared to those previously introduced in [17], where an allocation technique assigns users to subbands and beams using a greedy method based on a gradual relaxation of the correlation constraints between channel gains. These methods will be referred to by OMA-CR and NOMA-CR (CR for correlation relaxation) in the performance analysis part. Note that the method in [17] was shown to significantly outperform well-known previous methods in the literature like [11] and [15], in terms of both system throughput and user fairness. That is why it is used as a reference method for comparison.

More specifically, in OMA-CR, the user priorities are defined similarly as in steps 5 to 10 of Algorithm 1 (OMA-CN). Then, at each stage of the allocation algorithm, the selected user is assigned its best subband unless the latter has reached its maximum load or if there exists an already assigned user with a channel correlation higher than a predefined threshold. This threshold is then gradually increased until all users are assigned to subbands. Furthermore, the algorithm is adapted to the NOMA case (NOMA-CR) by defining another threshold for the correlation level between paired users on the same beam, and this threshold is also gradually relaxed until all users have been completely tested for possible pairing.

Algorithm 1: OMA-CN: OMA with Condition Number Criterion

Input : M, K, S, $\mathbf{h}_{k,s}, k = 1, ..., K, s = 1, ..., S$
Output: Assignment matrix $\mathbf{A}_{\text{OMA-CN}}$ of size $S \times M$

1 $\mathbf{A}_{\text{OMA-CN}} = \mathbf{0}$;
2 $n_s = 0$, $s = 1, \ldots, S$; // Variable containing the number of beams on subband s
3 $\mathbf{H}_s = []$, $s = 1, \ldots, S$; // Channel matrix of subband s
4 $\mathbf{E}(s,k) = \|\mathbf{h}_{k,s}\|$; // $\mathbf{E}$ is the $S \times K$ matrix containing the users channel norms on each subband
5 **for** $k = 1 : K$ **do**
6 $e_k^{\max} = \max_{s \in \mathcal{S}} \mathbf{E}(s,k)$;
7 $e_k^{\text{second}} = \max_{s \in \mathcal{S}, \mathbf{E}(s,k) \neq e_k^{\max}} \mathbf{E}(s,k)$;
8 $d_k = e_k^{\max} - e_k^{\text{second}}$;
9 **end for**
10 Sort the vector $\mathbf{d} = [d_1, \cdots, d_K]$ by descending order; $\mathbf{k_d}$ is the resulting vector of sorted user indices ; // $\mathbf{k_d}(1)$ is the index of the user with the highest priority
11 **for** $k = \mathbf{k_d}(1) : \mathbf{k_d}(S)$ **do**
12 $s^* = \text{argmax}_{s \in \mathcal{S}, n_s = 0} \mathbf{E}(s,k)$;
13 $n_{s^*} = n_{s^*} + 1$; $\mathbf{A}_{\text{OMA-CN}}(s^*, n_{s^*}) = k$;
14 $\mathbf{H}_{s^*} = [\mathbf{h}_{k,s^*}]$;
15 **end for**
16 **for** $k = \mathbf{k_d}(S+1) : \mathbf{k_d}(\textit{length}(\mathbf{k_d}))$ **do**
17 **for** $s = 1 : S$ **do**
18 **if** $n_s < M$ **then**
19 $\mathbf{H}_s^{\text{test}} = [\mathbf{H}_s; \mathbf{h}_{s,k}]$;
20 $CN_s^{\text{test}} = \text{cond}(\mathbf{H}_s^{\text{test}} (\mathbf{H}_s^{\text{test}})^H)$;
21 **else**
22 $CN_s^{\text{test}} = \infty$;
23 **end if**
24 **end for**
25 $s^* = \text{argmin}_s CN_s^{\text{test}}$;
26 $\mathbf{H}_{s^*} = \mathbf{H}_{s^*}^{\text{test}}$;
27 $n_{s^*} = n_{s^*} + 1$; $\mathbf{A}_{\text{OMA-CN}}(s^*, n_{s^*}) = k$;
28 **end for**

E. Adaptation to multi-antenna receivers

We move now to the adaptation of Algorithms 1, 2, and 3 to the case of multi-antenna receivers. Recall that vectors $\mathbf{h}_{k,s}$ become $N \times M$ matrices, which are now used in steps 4 to 9 of Algorithm 1. Then, in steps 14 and 19, vectors $\mathbf{h}_{k,s}$ are replaced by $\mathbf{z}_{k,s}\mathbf{h}_{k,s}$, with $\mathbf{z}_{k,s} = \frac{1}{\sqrt{N}}[1\,1\,\cdots\,1]^T$. The same goes for step 9 in Algorithm 2 and steps 8 and 45 in Algorithm 3. However, rate calculations in Algorithms 2 and 3 now rely on the combining methods explained in Section II and using Eqs. (13), (14) and (15).

F. Complexity analysis

In this part, we mainly focus on the single-antenna receiver case to derive the complexity of the different proposed allocation techniques. However, the main analysis and conclusions also hold for the case of multi-antenna users. The complexity is evaluated in terms of the number of complex multiplications (CMs).

Starting with the OMA-CR method [17], which relies on the correlation coefficient criterion: each correlation evaluation costs M CMs. For a chosen user k, $k-1$ correlations must be evaluated, one with every already assigned user on each of the available subbands. This leads to a maximum number of CMs by the order of $MN_{\text{it}}\sum_{k=1}^{K}(k-1) \approx K^2MN_{\text{it}}/2$, where N_{it} is the average number of correlation relaxation iterations, generally around 10. For the case of a full-load transmission, i.e., $K = SM$, the total is by the order of $M^3S^2N_{\text{it}}/2$.

Consider now the OMA-TR method. Its computational load is mainly dominated by the precoding step that involves matrix multiplication and inversion. Suppose first that it is applied without the iterative inversion technique and that the inverse of a square matrix of size $n \times n$ requires n^3 CMs. When testing the assignment of the k^{th} user on each one of the S subbands, we can consider an average number of used beams per subband: $n_k = (k-1)/S$. Based on Eq. (3), step 10 of Algorithm 2 costs around $2Mn_k^2 + n_k^3$ CMs. By taking $n_k \approx k/S$, this yields an approximate total of $\sum_{k=1}^{K}(2Mn_k^2 + n_k^3)S \approx M^4S^2$ when $K = SM$.

When applying the iterative inversion within OMA-TR, by inspecting Eq. (19), we can verify that it necessitates i^3+5i^2+Mi+M CMs at the $(i+1)^{th}$ iteration. At the same time, the direct calculation requires $M(i+1)^2+(i+1)^3 = i^3+(M+3)i^2+(2M+3)i+M$ CMs. Consequently, a gain in complexity by the order of $S\sum_{k=1}^{K-1}\left((M-2)\left(\frac{k}{S}\right)^2+(M+3)\frac{k}{S}\right) \approx \frac{MK^3}{3S} + \frac{MK^2}{2}$ CMs is obtained by applying the iterative inversion, yielding $\frac{M^4S^2}{3} + \frac{M^3S^2}{2}$ when $K = SM$. This corresponds to a reduction of around 33% of the complexity.

In the OMA-CN method, each CN calculation, at the k^{th} iteration, costs $Mn_k^2 + n_k^3$ CMs. Therefore, the approximate total cost of this method is $0.5M^4S^2$.

When comparing all three OMA methods, it appears that the OMA-CN allows the best tradeoff in complexity between OMA-CR and OMA-TR.

We move now to the estimation of the increase in computational load incurred by NOMA with respect to any of the OMA methods. As can be seen in Algorithm 3, in the iterative process, half of the users are re-explored for potential pairing with other users. To simplify the analysis, suppose that, at this point, each subband has $M/2$ allocated beams (in reality, this number may differ from one subband to another), with a total of $K/2 = MS/2$ allocated users for the full-load context. Also, consider the worst-case scenario where the selected user k is always the user with the highest channel gain in step 23 and that all users end up being paired (i.e., the test in step 49 is always False). Therefore, the k^{th} user is generally tested with $MS/2+1-k$ candidate users for potential pairing. This is also the number of times step 26 is repeated for user k. Therefore, the additional complexity of NOMA, which is mainly dominated by the precoding cost in step 26, is approximately $\sum_{k=1}^{K/2}(MS/2+1-k)(2M(M/2)^2+(M/2)^3) \approx 5/64M^5S^2$. In reality, a large proportion of the re-allocated users are paired as second users, therefore not necessitating the recalculation of the precoding matrix (steps 37 and 38 are used instead).

IV. Description of the channel measurement campaign

An intensive measurement campaign has been conducted on the Cité Scientifique Campus of the University of Lille (France). The MaMIMOSA radio channel sounder [24], [25] was used to collect 64×8 massive MIMO propagation channels at 325 outdoor and indoor positions. More precisely, an 8×8 vertical planar patch antenna array was deployed at the transmitter (Tx) (see Fig. 1(a)). The receiver (Rx) was equipped with a linear horizontal antenna array consisting of 8 EM-6116 omnidirectional antennas (see Fig. 1(b)).

The channel sounder operates at 5.89 GHz with 80 MHz bandwidth. Before transmission, data are modulated using orthogonal frequency-division multiplexing (OFDM) with 8192 subcarriers using a 12.21 kHz subcarrier spacing. At each transmit antenna, one pilot subcarrier out of eight consecutive subcarriers is used for measurement. This pilot transmission scheme was chosen at the Tx antennas so as to yield a frequency spacing between two consecutive measured subcarriers on a single antenna of 97.66 kHz. The transmitter is placed on a 10 m high platform on a four-storey building of the campus. The indoor environment corresponds to the interior of 3 floors of a facing building and the ground floor of two other buildings. Their interior consists of office and laboratory rooms. Outdoor positions are randomly selected within a 300 m radius with respect to Tx. This setup allows the emulation of a crowded mobile system. The detailed description of the measurement campaign is provided in a separate measurement document [26], along with the Tx and different Rx positions on the campus and building maps.

As an illustrative example, Fig. 3 displays the channel gain for a specific link between an antenna from the Tx array and an antenna from the Rx array at five Rx positions selected from the 325 measured positions, as specified in Fig. 2. Position 145, located in a building with a window facing the Tx, represents a line of sight channel. Position 139, also located in the same building, has a lower gain due to the Rx not being in front of a window. Positions 250 and 277 are outdoor, with

position 250 being more critical due to a building separating it from the Tx. Finally, position 315 is in a more open area but farther from the Tx. The channel gains can exhibit significant variations between the positions, with a maximum difference of 45 dB observed in this scenario. The effect of this variability of the channel gain on users' throughput will be discussed in Section V.

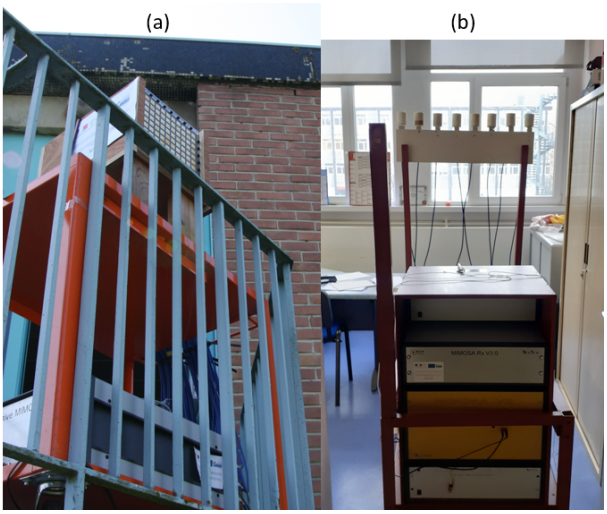

Fig. 1: (a) Tx (b) Rx

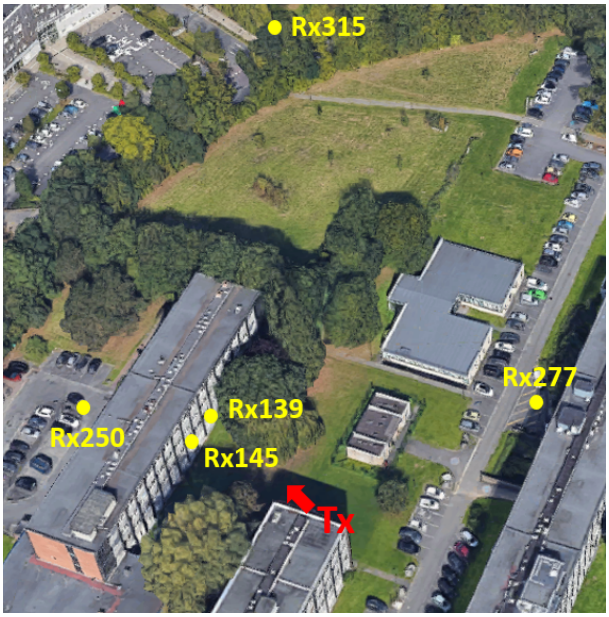

Fig. 2: Top-view of the measurement campaign on the Lille University campus with the Tx position and 5 Rx positions selected from the 325 measured positions

V. Results and Discussions

In order to evaluate the performance of the proposed algorithms, a set of system configurations was generated. Each configuration corresponds to a subset of K randomly selected users from a total set of 325 measured user positions. The set of 818 measured subcarriers was utilized to create 409

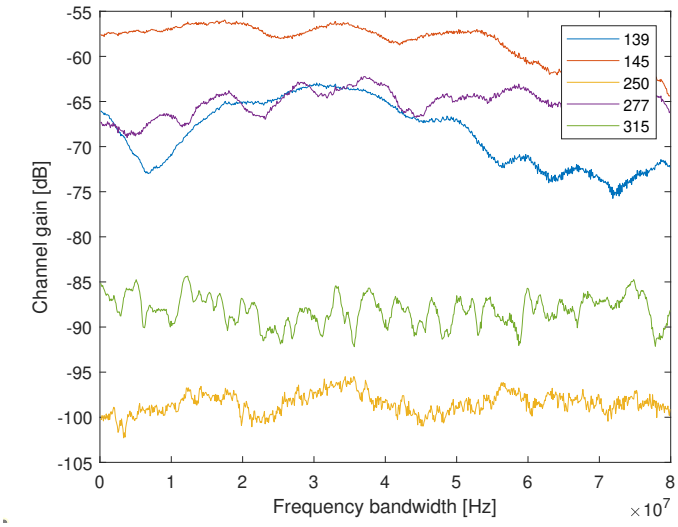

Fig. 3: Channel transfer functions obtained for 5 Rx positions

subbands by averaging every two consecutive subcarriers, resulting in subbands with a bandwidth of 195.32 kHz, which is similar to that of a resource block in the 5G standard. Furthermore, a subset of $S = 5$ was randomly chosen for each user setup. The measured channel matrices, as described in Section IV, are used as the matrices $\mathbf{H}$ in algorithms 1, 2, and 3. In the subsequent analysis, $M = 64$ transmit antennas were considered, and the number of active users was first set to $K = 320$, which corresponds to a full-load usage of spatial resources in OMA ($K = SM$). The intra-beam power allocation coefficient, $\alpha_{n,s,1}$, was fixed at 0.3. To begin, the study focused on single antenna receivers, i.e., $N = 1$. The system parameters are summarized in Table I.

TABLE I: Summary of the system parameters

S	5
M	64
N	1-8
K	80-320
P_{tot}	0.1-10 W

In Fig. 4, the total system spectral efficiency (SE) and user fairness are evaluated in terms of the base station power budget, P_{tot}. The results demonstrate the superiority of NOMA over OMA, with a gain in SE of approximately 600 bps/Hz observed at P_{tot} = 10 W between NOMA-TR and OMA-TR. Furthermore, it can be observed that the CN criterion yields performance that is very close to the total SE and superior to the correlation criterion in both the OMA and NOMA setups. This observation confirms that when used within a well-designed resource allocation method, the CN proves to be a good representative of the system rate performance. Furthermore, this result comes with a quasi-equality between NOMA-CN and NOMA-TR fairness, as shown in Fig. 4(b).

A deeper analysis of the obtained results proves that the NOMA-TR and NOMA-CN strategies allow better conditioning of the users' channel matrix when compared to NOMA-CR and all three OMA methods by significantly reducing the

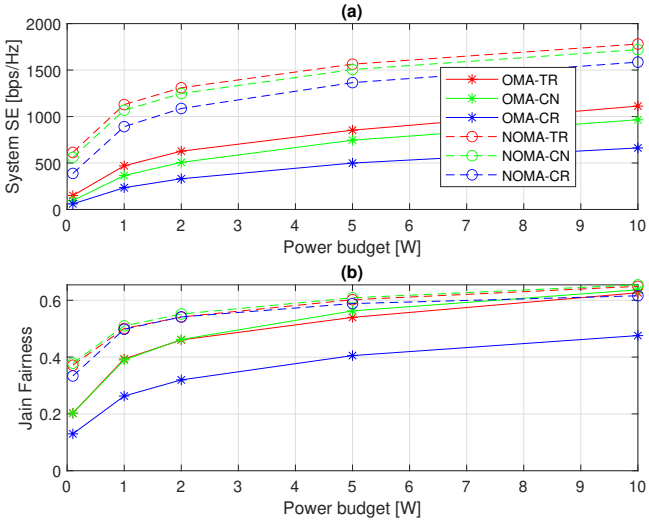

Fig. 4: System spectral efficiency (a) and Jain fairness (b) vs. P_{tot} in Watts, for single-antenna receivers

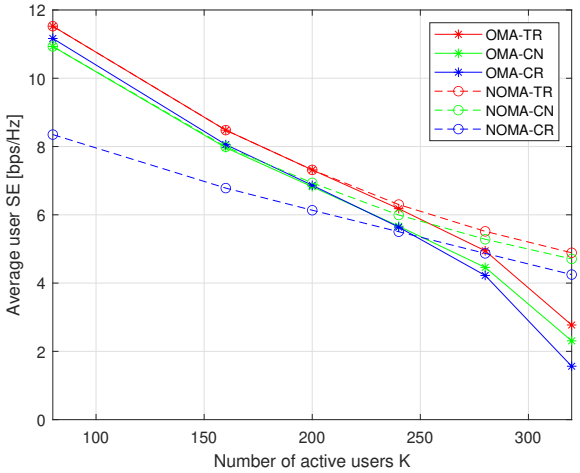

Fig. 5: Average user spectral efficiency vs. the number of active users K, for single-antenna receivers

number of beams per subband necessary to accommodate all active users. Indeed, the average number of beams per subband is 46 and 45 on average for NOMA-TR and NOMA-CN, respectively, compared to 64 for OMA.

In order to further evaluate the performance of the proposed approaches, Fig. 5 presents the average user SE obtained by the different methods when the number of active users is smaller or equal to SM, for a fixed base station power budget of P_{tot} = 5W. This scenario represents a situation where a full loading of the available spatial resources in OMA is often not required, and the number of beams used per subband decreases as the number of active users decreases. The results indicate that at high values of K, NOMA is the best strategy for any of the three allocation criteria since it reduces the number of beams by pairing some of the users when necessary. However, when the number of users is lower than 240, NOMA does not offer significant gains compared to OMA when the TR and CN criteria are used. This is because, in such conditions, OMA-TR requires a lower number of beams per subband than 48, while NOMA only reduces this number to around 43. Therefore, in a non-critical regime that only requires a moderate number of beams, further reducing the number of beams by NOMA is not necessary. Additionally, in this operating region, the CR criterion even leads to worse performance with NOMA compared to OMA, which highlights its suitability for congested systems only. For low values of K, the correlation relaxation mechanism sometimes leads to non-efficient pairings that are mostly avoided by the other criteria.

Fig. 6 shows the average system SE and Jain fairness of NOMA-CN applied for multi-antenna receivers with the combining techniques presented in Section II, in terms of N, for M = 64, K = 320 and $P_{\text{tot}} = 5W$. Our proposed IBI cancelation combining technique is referred to as "NOMA-CN-IBI-comb", and our MMSE combining technique to as "NOMA-CN-MMSE-comb". The reference method introduced in [13] and the EGC combining are respectively called "NOMA-CN-ref-comb" and "NOMA-CN-mean-comb". Our two proposed detection methods present very close performance and significantly outperform the reference method as well as the average detection in both system SE and user fairness. The gain is especially significant for second users since the implementation of the interference cancellation technique is only beneficial to second users' receivers, whereas interference between first users is canceled by the precoding step. This observation is further confirmed in Fig. 7, which represents the average individual rates of second users in terms of N. One can see how the performance of the proposed detection increases with N, while those of the other two methods are relatively insensitive to the increase in receive spatial diversity. This means that our detection methods better exploit this additional diversity. However, the individual second users' rates are globally low since these users generally have poor channel conditions.

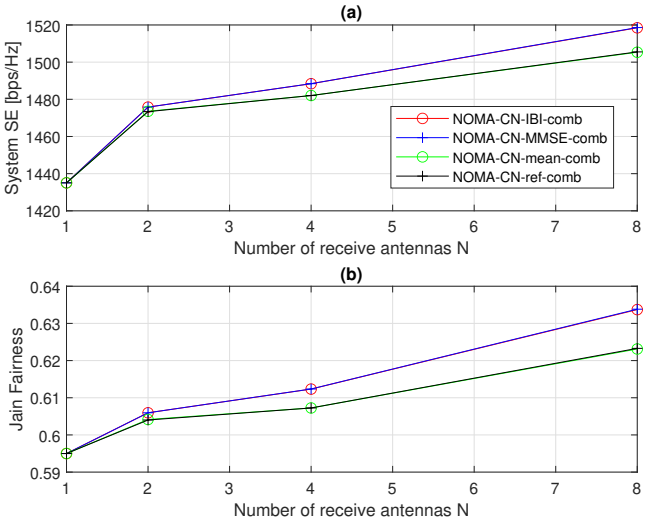

Fig. 6: System spectral efficiency (a) and Jain fairness (b) vs. the number of receive antennas N, for multi-antenna receivers

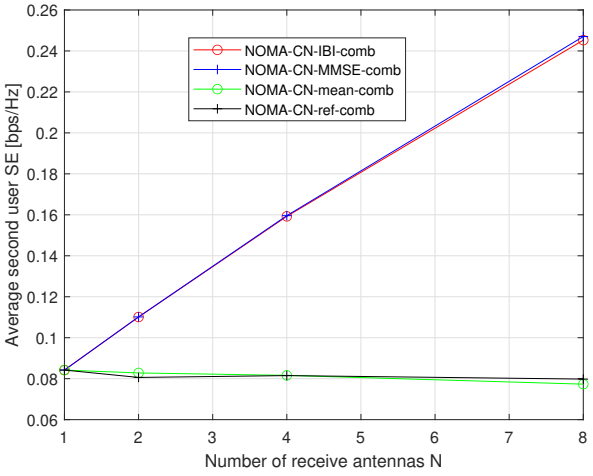

Fig. 7: Average spectral efficiency of second users vs. the number of receive antennas N, for multi-antenna receivers

VI. Conclusion

This paper presents new user pairing criteria and resource allocation techniques for congested massive MIMO-NOMA systems. These methods enable better exploitation of transmit antenna diversity compared to previous pairing methods that rely on the correlation criterion. Our results demonstrate that the use of the condition number criterion yields performance that is very close to the optimal rate criterion, whereas the correlation criterion sometimes leads to disadvantageous pairing situations. The performance gain of the condition number criterion over the correlation criterion comes with a moderate increase in complexity, which can be further reduced by an iterative matrix inversion technique based on the bordering method. The proposed allocation techniques are also adapted to the case of multi-antenna receivers, where we introduce two enhanced signal detection methods that provide better exploitation of receive spatial diversity compared to a well-referenced method in the literature.

References

[1] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO Linear Precoding: A Survey," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3920–3931, 2018.

[2] R. Chataut and R. Akl, "Massive MIMO systems for 5G and beyond networks—overview, recent trends, challenges, and future research direction," *Sensors*, vol. 20, no. 10, p. 2753, 2020.

[3] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2016.

[4] J. Flordelis, F. Rusek, X. Gao, G. Dahman, O. Edfors, and F. Tufvesson, "Spatial separation of closely-located users in measured massive MIMO channels," *IEEE Access*, vol. 6, pp. 40 253–40 266, 2018.

[5] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[6] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non orthogonal multiple access for 5G and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[7] S. Özyurt, E. P. Simon, and J. Farah, "NOMA with zero-forcing V-BLAST," *IEEE Communications Letters*, vol. 24, no. 9, pp. 2070–2074, 2020.

[8] J. Farah, E. P. Simon, P. Laly, and G. Delbarre, "Efficient combinations of NOMA With distributed antenna systems based on channel measurements for mitigating jamming attacks," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2212–2221, 2021.

[9] J. Farah, J. Akiki, and E. P. Simon, "Energy-efficient techniques for combating the influence of reactive jamming using non-orthogonal multiple access and distributed antenna systems," in *2019 Wireless Telecommunications Symposium (WTS)*, 2019, pp. 1–7.

[10] J. Farah, A. Kilzi, C. A. Nour, and C. Douillard, "Power minimization in distributed antenna systems using non-orthogonal multiple access and mutual successive interference cancellation," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 873–11 885, 2018.

[11] S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE access*, vol. 5, pp. 565–577, 2016.

[12] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4438–4454, 2016.

[13] X. Chen, F.-K. Gong, G. Li, H. Zhang, and P. Song, "User pairing and pair scheduling in massive MIMO-NOMA systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 788–791, 2018.

[14] J. Wang, Y. Li, C. Ji, Q. Sun, S. Jin, and T. Q. S. Quek, "Location-based MIMO-NOMA: multiple access regions and low-complexity user pairing," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2293–2307, 2020.

[15] B. Kim, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Military Communications Conference, MILCOM*. IEEE, 2013, pp. 1278–1283.

[16] E. P. Simon, J. Farah, and P. Laly, "Performance evaluation of massive MIMO with beamforming and nonorthogonal multiple access based on practical channel measurements," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 6, pp. 1263–1267, 2019.

[17] E. P. Simon, J. Farah, P. Laly, and G. Delbarre, "A gradual resource allocation technique for massive mimo-noma," *IEEE Antennas and Wireless Propagation Letters*, vol. 21, no. 3, pp. 476–480, 2021.

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

[19] J. Farah, E. Sfeir, C. A. Nour, and C. Douillard, "New resource allocation techniques for base station power reduction in orthogonal and non-orthogonal multiplexing systems," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 618–624.

[20] S. K. Mohammed and E. G. Larsson, "Improving the performance of the zero-forcing multiuser MISO downlink precoder through user grouping," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 811–826, 2015.

[21] D. K. Faddeev and V. N. Faddeeva, "Computational methods of linear algebra," *Zapiski Nauchnykh Seminarov POMI*, vol. 54, pp. 3–228, 1975.

[22] P. W. Wolniansky, G. J. Foschini, G. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Signals, Systems, and Electronics*, 1998.

[23] M.-J. Youssef, J. Farah, C. A. Nour, and C. Douillard, "Waterfilling-based resource allocation techniques in downlink non-orthogonal multiple access (NOMA) with single-user MIMO," in *IEEE Symposium on Computers and Communications (ISCC)*, 2017, pp. 499–506.

[24] D. Gaillot, P. Laly, N. Dahmouni, G. Delbarre, M. Van den Bossche, G. Vermeeren, E. Tanghe, E. Simon, W. Joseph, L. Martens *et al.*, "Measurement of the V2I massive radio channel with the MaMIMOSA sounder in a suburban environment," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2021, pp. 1–4.

[25] P. Laly, D. P. Gaillot, G. Delbarre, M. Van den Bossche, G. Vermeeren, F. Challita, E. Tanghe, E. P. Simon, W. Joseph, L. Martens *et al.*, "Massive radio channel sounder architecture for 5G mobility scenarios: MaMIMOSA," in *14th IEEE European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–5.

[26] E. P. Simon, J. Farah, and P. Laly, "Report on the 2023 massive MIMO channel measurement campaign," University of Lille, Tech. Rep., 2023. [Online]. Available: https://hal.science/hal-03941185

Algorithm 3: NOMA-CN and NOMA-TR

Input : M, K, S, $\mathbf{h}_{k,s}, k = 1, ..., K, s = 1, ..., S$, P_{tot}, $\mathbf{A}_{\text{OMA}} = \mathbf{A}_{\text{OMA-CN}}$ or $\mathbf{A}_{\text{OMA-TR}}$, n_s, $s = 1, \ldots, S$, $\mathbf{k_d}$

Output: Assignment matrices $\mathbf{A}^{(1)}_{\text{NOMA}}$, $\mathbf{A}^{(2)}_{\text{NOMA}}$ of size $S \times M$ for strong and weak users respectively

// **Initialization:**
$\mathbf{A}^{(1)}_{\text{NOMA}} = \mathbf{0}$; $\mathbf{A}^{(2)}_{\text{NOMA}} = \mathbf{0}$;
$\mathbf{u} = \mathbf{k_d}(\text{length}(\mathbf{k_d})/2 + 1 : \text{length}(\mathbf{k_d}))$; // $\mathbf{u}$ vector containing the last $K/2$ users allocated in the OMA phase
$\mathbf{v} = \mathbf{k_d}(1 : \text{length}(\mathbf{k_d})/2)$; // $\mathbf{v}$ vector containing the first $K/2$ users allocated in the OMA phase
$\mathbf{A}^{(1)}_{\text{NOMA}}$ is initialized from $\mathbf{A}_{\text{OMA}}$ for users $\in \mathbf{v}$;
$n_s^{(1)}$ is the number of beams with allocated users on beam s for $\mathbf{A}^{(1)}_{\text{NOMA}}$;
for $s = 1 : S$ **do**
  $\mathbf{H}_s = [\mathbf{h}_{s,\mathbf{A}^{(1)}_{\text{NOMA}}(s,1)}; \mathbf{h}_{s,\mathbf{A}^{(1)}_{\text{NOMA}}(s,2)}; ...; \mathbf{h}_{s,\mathbf{A}^{(1)}_{\text{NOMA}}(s,n_s^{(1)})}]$;
  Calculate the precoding matrix $\mathbf{W}_s$ for matrix $\mathbf{H}_s$ using Eqs. (3) and (4);
  for $j = 1 : n_s^{(1)}$ **do**
    // Calculate the rate for user allocation $\mathbf{A}^{(1)}_{\text{NOMA}}$ (no second users yet) and $\mathbf{W}_s$
    $R_{j,s}$ is found by Eq. (7)
  end for
  $R_s = \sum_{j=1}^{n_s^{(1)}} R_{j,s}$;
end for
// **Iterative Process:**
for $k = \mathbf{u}(1) : \mathbf{u}(\textit{length}(\mathbf{u}))$ **do**
  for $s = 1 : S$ **do**
    for $i = 1 : n_s^{(1)}$ **do**
      if $\mathbf{A}^{(2)}_{\textit{NOMA}}(s,i) == 0$ **then**
        // no 2nd user yet on beam i
        $\mathbf{A}^{(2)\text{test}}_{\text{NOMA}} = \mathbf{A}^{(2)}_{\text{NOMA}}$; $\mathbf{H}^{\text{test}}_s = \mathbf{H}_s$;
        if $\|\mathbf{h}_{k,s}\| > \|\mathbf{h}_{\mathbf{A}^{(1)}_{\textit{NOMA}}(s,i),s}\|$ **then**
          $\mathbf{A}^{(2)\text{test}}_{\text{NOMA}}(s,i) = \mathbf{A}^{(1)}_{\text{NOMA}}(s,i)$; // k is the strong user in the pair $(k, \mathbf{A}^{(1)}_{\text{NOMA}}(s,i))$
          $\mathbf{H}^{\text{test}}_s(i,:) = \mathbf{h}_{s,k}$;
          Calculate $\mathbf{W}^{\text{test}}_s$ for matrix $\mathbf{H}^{\text{test}}_s$ using Eqs. (3) and (4);
          for $j = 1 : n_s^{(1)}$ **do**
            // Calculate the rate using $\mathbf{W}^{\text{test}}_s$, $\mathbf{H}^{\text{test}}_s$ and $\mathbf{A}^{(2)\text{test}}_{\text{NOMA}}$
            if $\mathbf{A}^{(2)}_{\textit{NOMA}}(s,j) == 0$ **then**
              find $R_{j,s}$ by Eq. (7);
            else
              $R_{j,s} = R_{j,s,1} + R_{j,s,2}$, with $R_{j,s,1}$ and $R_{j,s,2}$ given by Eqs. (5) and (6);
            end if
            $R^{\text{tot}}_{(s,i)} = \sum_{j=1}^{n_s^{(1)}} R_{j,s} + \sum_{s' \neq s} R_{s'}$;
          end for
        else
          $\mathbf{A}^{(2)\text{test}}_{\text{NOMA}}(s,i) = k$; // k is the weak user in the pair $(k, \mathbf{A}^{(1)}_{\text{NOMA}}(s,i))$
          Calculate the rate using $\mathbf{W}_s$, $\mathbf{H}^{\text{test}}_s$ and $\mathbf{A}^{(2)\text{test}}_{\text{NOMA}}$ by repeating steps 27 to 35;
        end if
      end if
    end for
  end for
  $(s^*, i^*) = \text{argmax}_{s,i} R^{\text{tot}}_{(s,i)}$; $R^* = R^{\text{tot}}_{(s^*,i^*)}$;
  // Compare the best rate to the case where k is assumed to be a single user on its initial subband s_k
  $[s_k, i_k] = \text{find}(\mathbf{A}_{\text{OMA}} == k)$; $\mathbf{H}^{\text{test}}_{s_k} = [\mathbf{H}_{s_k}; \mathbf{h}_{s_k,k}]$;
  Calculate $\mathbf{W}^{\text{test}}_{s_k}$ for matrix $\mathbf{H}^{\text{test}}_{s_k}$; // No need to recompute $\mathbf{W}_s$ for $s \neq s_k$
  Calculate R^{test}_s by repeating steps 27 to 35, for $s = 1, \cdots, S$;
  Calculate $R^{\text{single}} = \sum_s R^{\text{test}}_s$;
  if $R^{\textit{single}} > R^*$ **then**
    $n^{(1)}_{s_k} = n^{(1)}_{s_k} + 1$; $\mathbf{A}^{(1)}_{\text{NOMA}}(s_k, n^{(1)}_{s_k}) = k$; Update $\mathbf{H}_{s_k}$, $\mathbf{W}_{s_k}$ and R_s for $s = 1, \cdots, S$;
  else
    $k' = A^{(1)}_{\text{NOMA}}(s^*, i^*)$;
    if $\|\mathbf{h}_{k,s^*}\| > \|\mathbf{h}_{\mathbf{k}',s^*}\|$ **then**
      $\mathbf{A}^{(2)}_{\text{NOMA}}(s^*, i^*) = k'$; $\mathbf{A}^{(1)}_{\text{NOMA}}(s^*, i^*) = k$; Update $\mathbf{H}_{s^*}$, $\mathbf{W}_{s^*}$ and R_{s^*};
    else
      $\mathbf{A}^{(2)}_{\text{NOMA}}(s^*, i^*) = k$; Update R_{s^*};
    end if
  end if
end for