On User Pairing in NOMA Uplink

Mohammad A. Sedaghat, and Ralf R. Müller, Senior Member, IEEE

Abstract

User pairing in Non-Orthogonal Multiple-Access (NOMA) uplink based on channel state information is investigated considering some predefined power allocation schemes. The base station divides the set of users into disjunct pairs and assigns the available resources to these pairs. The combinatorial problem of user pairing to achieve the maximum sum rate is analyzed in the large system limit for various scenarios, and some optimum and sub-optimum algorithms with a polynomial-time complexity are proposed. In the first scenario, $2M$ users and the base station have a single-antenna and communicate over $M$ subcarriers. The performance of optimum pairing is derived for $M \to \infty$ and shown to be superior to random pairing and orthogonal multiple access techniques. In the second setting, a novel NOMA scheme for a multi-antenna base station and single carrier communication is proposed. In this case, the users need not be aware of the pairing strategy. Furthermore, the proposed NOMA scheme is generalized to multi-antenna users. It is shown that random and optimum user pairing perform similarly in the large system limit, but optimum pairing is significantly better in finite dimensions. It is shown that the proposed NOMA scheme outperforms a previously proposed NOMA scheme with signal alignment.

I. INTRODUCTION

Non-Orthogonal Multiple-Access (NOMA) has been introduced recently as one of the key technologies for the next generation of wireless networks referred to as 5G [1]–[3]. Compared to Orthogonal Multiple-Access (OMA) techniques, NOMA allows a larger number of users to be connected to a wireless network simultaneously [4]. Furthermore, it has been shown that NOMA achieves a larger total throughput in both uplink and downlink of a wireless network compared to OMA techniques [5], [6]. Compared to conventional opportunistic schemes, a higher fairness is achieved by NOMA [4], [7]. The low latency property of NOMA together with realizing a
tradeoff between total throughput and fairness makes NOMA very attractive for 5G and a serious competitor for well-known Orthogonal Frequency Division Multiple-Access (OFDMA) [1], [8].

NOMA can be done in power domain by allowing more than one users to send over a resource block [9] and also in code domain by assigning different sparse codes to users, known as Sparse Code Multiple-Access (SCMA) [10], [11]. This paper is about power domain NOMA, therefore in the remaining parts of this paper, by NOMA we refer to power domain NOMA.

The idea of NOMA is based on the fact that users normally have different channel qualities and rate demands. Information theory shows that orthogonal signaling techniques are not optimum in such cases [12]–[14]. In NOMA, base stations share each resource among some users with different channel qualities, and the receiver in a NOMA system uses Successive Interference Cancellation (SIC) to detect the transmitted signals. It is known from information theory that, for a given set of rates and users with different channel qualities, OMA methods are suboptimum with respect to required total transmit power on multiple-access channels [12], [13]. Therefore, NOMA methods can be more power efficient than OMA methods when users with different channel qualities are grouped. Different channel qualities of the users typically result from fading and the near-far effect in cellular networks.

NOMA with randomly deployed mobile stations in downlink is discussed in [2], and it is shown that NOMA achieves higher ergodic sum rates compared to OMA schemes. Impact of user pairing in NOMA downlink is investigated in [3]. Cooperative NOMA in downlink is introduced in [15] in which after the direct transmission phase, users cooperate via short range communication channels. Fairness in NOMA is investigated in [7] by studying some power allocation techniques based on instantaneous and statistical channel state information. Joint user pairing and power allocation in a multi-carrier downlink channel with single-antenna users and base stations is addressed numerically in [16] using some optimization techniques. The authors in [5] provide a mathematical proof for the superiority of NOMA in downlink channel compared to OMA schemes from an optimization point of view. Furthermore, in [17], an efficient user pairing and power allocation scheme with low complexity is proposed for NOMA downlink.

In NOMA downlink, SIC should be implemented at user terminals which needs high processing power especially when the number of active users at each resource block is large. One should also consider the loss of practical SIC for large number of users [18]. To this end, grouping only two users at each resource block in NOMA downlink has become more popular.
In uplink, SIC is at base stations, therefore users need not be aware of the modulation and coding schemes employed by the other users. Furthermore, base stations have enough processing power to perform SIC. NOMA can also allow users to transmit in uplink in a grant-free manner which reduces latency significantly.

NOMA uplink using a sub-optimal multi-user detection technique is investigated in [19], and an iterative detection and decoding scheme is proposed in [20] which improves the system performance significantly. Furthermore, an iterative method for joint subcarrier and power allocation is proposed in [20].

Along with single-antenna applications, NOMA has been also proposed for multi-antenna wireless networks. Ergodic capacity of Multiple-Input Multiple-Output (MIMO) NOMA is discussed in [21] by studying optimum power allocation. The authors in [22] propose a new MIMO NOMA scheme by designing the precoding and detection matrices. In [22], the number of antennas at each user terminal should be larger than the number of antennas at base stations. A more general MIMO NOMA is introduced in [23] based on signal alignment for a multi-user MIMO system in which both the users and the base station are equipped with antenna arrays. It has been shown that the number of supported users in a single frequency band can be up to twice the number of antennas at the base station under some certain conditions [23]. The proposed method in [23] is applicable to both downlink and uplink. A NOMA scheme for massive MIMO with limited feedback is proposed in [24] in which a massive MIMO NOMA channel is decomposed into multiple separated single-input single-output NOMA channels.

Although, user pairing and power allocation for NOMA uplink have been studied previously, to the best of our knowledge there is no analytical performance evaluation for optimum user pairing in NOMA uplink. In this paper, we answer the question how much performance improves if the users in NOMA uplink are paired optimally. Furthermore, we show that optimal user pairing in uplink can be implemented using some algorithms with polynomial time complexity. To this end, we consider the general problem of user pairing in NOMA uplink in a cellular network with some fixed power allocation schemes. We consider the investigation of optimum joint power allocation and user pairing as a future work. At each cell, a central base station assigns the available resources including some subcarriers and spatial dimensions to some users to achieve the best performance in the sense of total rate of the users. The inter-cell interference is neglected and a single isolated cell is considered. The user terminals and the base station in the considered
cell may all have multi-antenna arrays. The base station first divides the users into some groups and lets the users at each group transmit over one of the subcarriers. Then, the signals at each subcarrier are detected by dividing the users, transmitting at each subcarrier, into some pairs and applying SIC to each pair.

Solving such a problem in general is very complicated. Thus, in this paper we consider various scenarios. The contributions of the paper for these scenarios are summarized as follows.

- **Single antenna multiple subcarriers (SAMS):** In the SAMS setting, the users and the base station are assumed to have a single-antenna. The users are divided into disjunct pairs and every subcarrier is assigned to a pair. A sub-optimum pairing algorithm with polynomial time complexity is proposed and shown to have almost the same performance as optimum method. Furthermore, a large system performance analysis of optimum pairing for the case that the users have independent and identically distributed (iid) Gaussian small scale fading coefficients at different subcarriers, is presented. It is shown that for $2M$ users and $M$ subcarriers, the total rate of the system divided by $M \log_2 \log M$ converges to 1 in probability as $M \to \infty$. Furthermore, for frequency-flat channels and large $M$, the performance of optimum pairing is derived and compared against the performance of random pairing and time division multiple-access (TDMA) at each subcarrier.

- **Multiple base station antennas single subcarrier (MBASS):**
  - The MBASS-SAU setting is the case of a multi-antenna base station and single-antenna users (SAU). We consider one single subcarrier and investigate the performance of optimum spatial dimension assignment. The base station divides the users into pairs and detects every pair using SIC. It is shown that the optimum pairing method has significant performance gain compared to random pairing in finite dimensions. We show that optimum assignment in this case can be implemented in polynomial time using the Hungarian algorithm with a modified cost matrix. Furthermore, a large system analysis of optimum pairing method is presented and it is shown that when the number of users and the number of antennas at the base station are large, the performances of both optimum and random pairing converge to the same limit.
  - Motivated by the work in [22], [23], the case of multi-antenna users (MAU) is considered as the MBASS-MAU setting. A new polynomial time NOMA scheme is proposed...
in which the users need not know the user pairing strategy and the base station pairs
the users based on channel state information. It is shown that, unlike to NOMA with
signal alignment [23], there are no requirements on the ranks of the channels and the
number of antennas at the user terminals. Furthermore, the users only need to know
their own channel. We show that the proposed method has better performance than
NOMA with signal alignment.

We use bold lowercase letters for vectors and bold uppercase letters for matrices. Conjugate
transpose of a matrix is denoted by $\cdot^\dagger$ and the transpose itself is shown by $\cdot^T$. The identity
matrix is shown by $I$. Moreover, the real and complex sets are shown by $\mathbb{R}$ and $\mathbb{C}$, respectively.
The determinant is denoted by $\det(\cdot)$, $\overset{p}{\Rightarrow}$ denotes the convergence in probability and $x \overset{p}{\leq} y$
implies that the random variable $x$ is less than or equal to $y$ with probability 1. Furthermore,$a \sim \mathcal{CN}(\mu, \sigma^2)$ means that $a$ is a complex Gaussian random variable with mean $\mu$ and variance $
igma^2$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A multi-user wireless cellular network with a central base station at each cell is considered.
For sake of analysis, we consider one individual cell and neglect the inter-cell interference. We
assume that the base station in the considered cell has $N$ antennas and there are $M$
subcarriers available for the communication between the users and the base station. Each subcarrier is
assigned to $K$ users simultaneously, thus the total number of users is $MK$. It is assumed that
each user has $L$ antennas.

Throughout the whole paper, we consider the uplink channel. It is assumed that the base
station is able to do one-step SIC which is explained in Definition I for a simple setting.

Definition 1: Let $x_1$ and $x_2$ be the iid Gaussian signals of two single-antenna users with
transmit power $P_1$ and $P_2$, respectively. Assume that a receiver with $N$ antennas knows the
codebooks of both users and the received signal at the receiver is

$$y = h_1 x_1 + h_2 x_2 + n,$$

for some known channel vectors $h_1, h_2$ and the additive white Gaussian noise $n \sim \mathcal{CN}(0, \sigma_n^2 I)$.
The receiver can first detect and decode the signal of one of the users using a Minimum Mean
Square Error (MMSE) detector and then cancel it from the received signal and detect the signal
of the other user. We call this well-known detection method one-step SIC throughout this paper. From information theory, the maximum total rate is obtained as $\log_2(1 + \frac{\|h_1\|^2 P_1 + \|h_2\|^2 P_2}{\sigma_n^2})$ \[25\].

In this paper, we limit the analysis to one-step SIC. One can consider SIC with higher steps at the expense of complexity. For given rates on the multiple-access channel with two users, the total transmit power is minimized if the base station first detects the user with the stronger channel and then the user with the weaker channel.

The fundamental problem of NOMA is how to assign the available resources to the users. In the considered network, the available resources are the subcarriers and the spatial dimensions obtained from the MIMO nature of the system. We analyze such a problem in the SAMS setting in Sections III and IV for frequency-selective and frequency-flat fading, respectively. In Sections V and VI we address the MBASS-SAU and MBASS-MAU setting, respectively.

In this paper a single cell uplink channel is considered in which several users are distributed within the cell around the base station. To model the path loss of the users, similar to \[23\], we assume that the users are uniformly distributed in a disc with radius $R_d$ and the base station is at the center of the disc. The shadowing effect is neglected in this paper. The small scale fading coefficients are assumed to be iid Gaussian distributed, and the path loss of $k$th user is modeled as $\frac{1}{r_0 + r_k^2}$ where $r_k$ is the distance between the $k$th user and the base station and $r_0$ is a constant to avoid singularity\[1\].

Although the proposed schemes in this paper can be applied to an uplink channel with any power allocation strategy, we limit the numerical and simulation results to two types of power control methods for the user terminals. One is when the users set their power proportional to the inverse of their path loss. In this case, the power multiplied by the path loss for all the users is equal to a constant. We call this method perfect power control (PPC). The second method is when the users transmit with the same power independent of their path loss which is called the equal power (EP) method in the sequel. Investigating optimum joint power allocation and user pairing in NOMA uplink analytically is an interesting future work.

\[1\]In practice, $r_0$ is on the order of a wavelength and accounts for the fact that free-space path loss does not apply in the near-field of antennas.
III. SAMS IN FREQUENCY-SELECTIVE FADING

In this section, we consider the problem of assigning each of the \( M \) subcarriers to two single-antenna users when the base station has a single antenna, i.e., \( K = 2 \), \( N = 1 \) and \( L = 1 \). Thus, there are \( 2M \) users in total. Assume without loss of generality that the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th user are selected to transmit on the \( m \)th subcarrier. The received signal at the base station for the \( m \)th subcarrier is

\[
y_m = \sqrt{d_{\ell_{2m-1}} h_{\ell_{2m-1},m} x_{\ell_{2m-1},m}} + \sqrt{d_{\ell_{2m}} h_{\ell_{2m},m} x_{\ell_{2m},m}} + n_m, \tag{2}
\]

where \( d_i, h_{i,j} \sim \mathcal{CN}(0, 1) \) and \( x_{i,j} \) model the path loss, the small scale fading and the transmitted signal of the \( i \)th user at the \( j \)th subcarrier, respectively, and \( n_m \) is additive white Gaussian noise with variance \( \sigma_n^2 \). It is assumed that the path loss of each user is identical at different subcarriers, but the fast fading coefficients differ. The signals of the users are assumed to be independent and Gaussian distributed. Let \( P_i \) be the transmit power of the \( i \)th user which is selected prior to the subcarrier assignment according to a power allocation method. Furthermore, it is assumed that the base station knows the \( d_i \)s and \( h_{i,j} \)s perfectly. Note that in general, power and subcarrier allocation can be done jointly. However, in this paper we assume that the power allocation has been done beforehand based on the path loss of the users.

At the beginning of each transmission interval, the base station determines the optimum pairing based on the channel information. We assume that during each transmission interval \( d_i \)s and \( h_{i,j} \)s are constant. Thus, the transmission interval must be shorter than the coherence time of the channel.

The base station applies one-step SIC to detect the signals with the sum rate

\[
R_m(\ell_{2m-1}, \ell_{2m}) \triangleq \log_2 \left( 1 + \frac{|h_{\ell_{2m-1},m}|^2 d_{\ell_{2m-1}} P_{\ell_{2m-1}} + |h_{\ell_{2m},m}|^2 d_{\ell_{2m}} P_{\ell_{2m}}}{\sigma_n^2} \right), \tag{3}
\]

at the \( m \)th subcarrier. For such a setting, the optimum user pairing is formulated as

\[
\Pi_{\text{opt}} = \arg\max_{(\ell_{2m-1}, \ell_{2m}) \in \Pi} \sum_{m=1}^{M} R_m(\ell_{2m-1}, \ell_{2m}), \tag{4}
\]

where \( \Pi \) models all the possible pairing sets. \( \Pi_{\text{opt}} \) denotes the optimal pairing scheme which gives the maximum sum rate.
A. A sub-optimum algorithm with polynomial time complexity

The total number of pairing sets is \((\frac{(2M)!}{2^M})\) which can be approximated using Stirling’s approximation as \(\sqrt{4\pi M} \left(\frac{\sqrt{2M}}{e}\right)^{2M}\). As observed, the total number of sets grows superexponential with \(M\), thus the optimum solution is not feasible to calculate even for moderate \(M\). To simplify the problem, we use the bounds provided in the following lemma.

**Lemma 1:** \(R_m\) is bounded by

\[
\frac{1}{2} \log_2 \left( 1 + 2 \frac{|h_{\ell_{2m-1},m}|^2 d_{\ell_{2m-1}} P_{\ell_{2m-1}}}{\sigma_n^2} \right) + \frac{1}{2} \log_2 \left( 1 + 2 \frac{|h_{\ell_{2m},m}|^2 d_{\ell_{2m}} P_{\ell_{2m}}}{\sigma_n^2} \right) \leq R_m(\ell_{2m-1}, \ell_{2m}) \leq \log_2 \left( 1 + \frac{|h_{\ell_{2m-1},m}|^2 d_{\ell_{2m-1}} P_{\ell_{2m-1}}}{\sigma_n^2} \right) + \log_2 \left( 1 + \frac{|h_{\ell_{2m},m}|^2 d_{\ell_{2m}} P_{\ell_{2m}}}{\sigma_n^2} \right). \tag{5}
\]

**Proof:** The lower bound can be easily proven using the inequality of arithmetic and geometric means. In fact, the lower bound is achievable using TDMA and the upper bound is the case that there is no interference between the users. The bounds become tight in the low Signal to Noise Ratio (SNR) regime. Assuming that the \(\ell_{2m-1}\)th and the \(\ell_{2m}\)th user transmit at the \(m\)th subcarrier, the main advantage of the bounds provided in Lemma [1] is that they follow the general form

\[
g_{\ell_{2m-1},m} + g_{\ell_{2m},m}, \tag{6}
\]

in which the contribution of the users in the sum rate can be decoupled. \(g_{i,j}\) approximately represents the contribution of \(j\)th user in the sum rate at the \(i\)th subcarrier. Using any of the bounds in Lemma [1] as an approximation for \(R_m(\ell_{2m-1}, \ell_{2m})\), we can solve the optimization problem in (4) in polynomial time according to the following lemma.

**Lemma 2:** Assume the pairing problem

\[
\max_{(\ell_{2m-1}, \ell_{2m}) \in \Pi} \sum_{m=1}^M g_{\ell_{2m-1},m} + g_{\ell_{2m},m}, \tag{7}
\]

where \(m \in \{1, \ldots, M\}, \ell_i \in \{1, \ldots, 2M\}\) and \(\Pi\) models all the pairing sets. The optimization problem can be solved optimally using the Hungarian algorithm with a polynomial time complexity.

**Proof:** To convert the problem to a minimization problem, we define

\[
\hat{g}_{i,j} \triangleq \max_{k,n} g_{k,n} - g_{i,j}. \tag{8}
\]
It can be easily shown that the solution pairs of the problem

$$\min_{(\ell_2, -1, \ell_2 m) \in \Pi} \sum_{m=1}^{M} \hat{g}_{2m-1, m} + \hat{g}_{2m, m},$$

(9)

are the same as the solution pairs of the original problem. Furthermore, \( \hat{g}_{i,j} \geq 0 \). Next, define \( Z \in \mathbb{R}^{2M \times M} \) such that its \((i, j)\)th entry is equal to \( \hat{g}_{i,j} \). To obtain the optimal pairs, we can solve an equivalent problem which is a linear assignment problem with the cost matrix \( Z_t \triangleq [Z Z] \).

The problem is the same problem as the job-worker problem in which we have \( 2M \) workers and \( M \) jobs and we want to assign each job to two workers. To solve such a problem, we copy the cost matrix \( Z \) for \( M \) new virtual jobs and minimize the cost by assigning each job to one worker exclusively. Finally, the workers assigned to the \( i \)th job and the \((i + M)\)th virtual job will be assigned to \( i \)th job which means two workers per job. This linear assignment problem can be solved optimally using the Hungarian algorithm with a polynomial time complexity [26].

The complexity of the Hungarian algorithm in this case is \( O(M^3) \) [27].

Due to the lack of space here we omit to explain the Hungarian algorithm. To see the details of the Hungarian algorithm, some simple examples and also an online demo video, please see [28]. Lemma 2 gives a sub-optimum algorithm for the user pairing problem. To derive the performance of this sub-optimum method, first we derive the optimum set to maximize the lower bound or the upper bound in Lemma 1, and then we calculate the sum rate for this given set.

B. A large system performance analysis for optimum pairing

Define \( \gamma_i \triangleq \frac{P_i d_i}{\sigma_n^2} \) which is the receive SNR of the \( i \)th user at the base station. Therefore,

$$R_m(\ell_2, -1, \ell_2 m) = \log_2 \left( 1 + \gamma_{\ell_2, -1} |h_{\ell_2, -1, m}|^2 + \gamma_{\ell_2, m} |h_{\ell_2, m}|^2 \right).$$

(10)

For the sake of analysis, we consider iid Gaussian \( h_{i,j} \) with variance 1. Furthermore, we assume that the users are distributed around the base station in a finite size area and all the users have nonzero \( \gamma_i \) at the base station. The following theorem states the main result of this subsection.

**Theorem 1:** Let \( h_{i,j} \) be iid Gaussian distributed with variance of 1 and \( c_1 \leq \gamma_i \leq c_2 \) for two finite constants \( c_1 \) and \( c_2 \). For \( M \rightarrow \infty \),

$$\max_{(\ell_2, -1, \ell_2 m) \in \Pi} \sum_{m=1}^{M} R_m(\ell_2, -1, \ell_2 m) \quad \frac{\log_2 \log M}{M} \rightarrow 1.$$ 

(11)

**Proof:** The proof is given in Appendix [A].

\[ \square \]
Based on the result of Theorem 1, it is easy to show that
\[
\lim_{M \to \infty} \mathbb{E} \max_{(\ell_{2m-1}, \ell_{2m}) \in \Pi} \sum_{m=1}^{M} R_m(\ell_{2m-1}, \ell_{2m}) / M \log_2 \log M = 1,
\] (12)
where the expectation \( \mathbb{E} \) is over the channel realizations.

Another interesting asymptotic behavior is the case of fixed \( M \) and \( \sigma_n^2 \to 0 \). One can use the two bounds presented in the proof of Theorem 1 and show that
\[
\max_{(\ell_{2m-1}, \ell_{2m}) \in \Pi} \sum_{m=1}^{M} R_m(\ell_{2m-1}, \ell_{2m}) / M \log_2(1 + c_3 \sigma_n^2 \log M) \xrightarrow{p} 1,
\] (13)
for \( \sigma_n^2 \to 0 \) and any arbitrary finite and positive constant \( c_3 \). We omit the proof here since it is quite similar to the proof of Theorem 1.

IV. SAMS IN FREQUENCY-FLAT FADING

If the channels of all users are flat in frequency, the channels are identical for all subcarriers, i.e., \( h_{i,j} = h_i \). We use the same setting as in Section III, i.e., \( N = L = 1 \) and \( K = 2 \). The goal is to obtain the best pairing set which maximizes the total rate
\[
\Pi_{\text{opt}} = \arg\max_{(\ell_{2m-1}, \ell_{2m}) \in \Pi} \sum_{m=1}^{M} R_m(\ell_{2m-1}, \ell_{2m}),
\] (14)
where in this case
\[
R_m \triangleq \log_2 \left( 1 + \frac{|h_{\ell_{2m-1}}|^2 d_{\ell_{2m-1}} P_{\ell_{2m-1}} + |h_{\ell_{2m}}|^2 d_{\ell_{2m}} P_{\ell_{2m}}}{\sigma_n^2} \right).
\] (15)
The following theorem presents the optimum user pairing strategy.

**Theorem 2:** If we sort the channel coefficients such that \( |h_1|^2 d_1 P_1 \geq |h_2|^2 d_2 P_2 \geq \cdots \geq |h_{2M}|^2 d_{2M} P_{2M} \), then the optimum pairing set which maximizes the total rate is
\[
\Pi_{\text{opt}} = \{(1, 2M), (2, 2M - 1), \ldots, (M, M + 1)\}.
\] (16)

**Proof:** Assume the contrary, i.e. \( \Pi = \{(1, 2M), (2, 2M - 1), \ldots, (M, M + 1)\} \) is not the optimum set. Then, there is at least one subset \( \{w_1, w_2, w_3, w_4\} \) where \( w_i \in \{1, \ldots, 2M\} \) in the optimum set that is paired as one of the following options
\[
\Pi_1 = \{(w_1, w_2), (w_3, w_4)\},
\] (17)
\[
\Pi_2 = \{(w_1, w_3), (w_2, w_4)\}.
\] (18)
It is easy to show that these two pairing schemes result in a lower total rate compared to the paring scheme \( \{(w_1, w_4), (w_2, w_3)\} \). This proves the theorem.

Next, we analyze the optimum pairing scheme for \( \gamma_i \triangleq \frac{d_i P_i}{\sigma_n^2} \) and \( h_i \sim \mathcal{CN}(0, 1) \) in the large system limit. In this case, we assume that the whole bandwidth including all the subcarriers is still smaller than the coherence bandwidth of the channel of the users such that the channel of each user is the same at different subcarriers. Let \( f(\gamma) \) be the empirical density function of the SNRs \( \gamma_i \).

**Theorem 3:** For \( h_i \sim \mathcal{CN}(0, 1) \), \( \gamma_i \sim f(\gamma_i) \) and \( M \to \infty \), the average rate of the users converges to

\[
\tilde{R} \triangleq \frac{1}{2M} \sum_{m=1}^{M} R_m \xrightarrow{p} \frac{1}{2} \log_2 \left( 1 + F^{-1}(t/2) + F^{-1}(1 - t/2) \right) dt,
\]

where

\[
F(z) = 1 - \int_{\mathbb{R}} e^{-z/x} f(x) dx,
\]

and \( F^{-1}(x) \) is the inverse of \( F(x) \) with respect to composition.

**Proof:** Let \( t_i \triangleq \gamma_i |h_i|^2 \) and assume that the samples are sorted as \( t_1 \geq t_2 \geq \cdots \geq t_{2M} \).

For large \( M \), \( t_i \) converges to \( F^{-1} \left( \frac{i-1}{2M} \right) \) in probability where \( F(\cdot) \) is the cumulative distribution function (cdf) of \( \gamma_i |h_i|^2 \) and \( F^{-1}(\cdot) \) is the inverse with respect to composition. For \( h_i \sim \mathcal{CN}(0, 1) \) and \( \gamma_i \sim f(\gamma_i) \), it can be shown that \( F(z) \) can be calculated from (20). Thus, for \( M \to \infty \)

\[
\tilde{R} = \frac{1}{2M} \sum_{m=1}^{M} R_m \xrightarrow{p} \frac{1}{2M} \sum_{m=1}^{M} \log_2 \left( 1 + t_m + t_{2M-m+1} \right)
\]

\[
= \frac{1}{2M} \sum_{m=1}^{M} \log_2 \left( 1 + F^{-1} \left( \frac{m-1}{2M} \right) + F^{-1} \left( 1 - \frac{m-1}{2M} \right) \right)
\]

\[
= \int_{0}^{1} \frac{1}{2} \log_2 \left( 1 + F^{-1}(t/2) + F^{-1}(1 - t/2) \right) dt.
\]

For the PPC strategy, i.e., \( \gamma_i = \gamma \), and \( M \to \infty \), we have

\[
\frac{1}{2M} \sum_{m=1}^{M} R_m \xrightarrow{p} \int_{0}^{1} \frac{1}{2} \log_2 \left( 1 + \gamma \log \left( \frac{1}{1-t/2} \right) + \gamma \log \left( \frac{2}{t} \right) \right) dt.
\]
As a benchmark, we also derive the average rate for random pairing. For large \( M \), the average rate of random pairing converges to
\[
\int_0^\infty \int_0^\infty \frac{1}{2} \log_2(1 + t_1 + t_2) g(t_1) g(t_2) dt_1 dt_2,
\]
(23)
where \( g(t) \) is the probability density function of \( t = \gamma_i|h_i|^2 \) which is given as
\[
g(t) = \frac{\partial}{\partial t} F(t) = \int_{\mathbb{R}} \frac{e^{-t/x}}{x} f(x) dx.
\]
(24)

Finally, in the following lemma, we prove that for the considered setting, the optimum pairing strategy for the max-min rate criterion is the same strategy as introduced in Theorem 2. The max-min rate criterion assures to maximize the minimum sum rate of the subcarriers.

**Theorem 4:** For \(|h_1|^2d_1P_1 \geq |h_2|^2d_2P_2 \geq \cdots \geq |h_{2M}|^2d_{2M}P_{2M}\), the optimum pairing set which is derived from
\[
\Pi_{\text{mm}} = \arg\max_{\Pi} \min_m R_m,
\]
(25)
is
\[
\Pi_{\text{mm}} = \{(1, 2M), (2, 2M - 1), \ldots, (M, M + 1)\}.
\]
(26)

**Proof:** The proof is similar to the proof of Theorem 2.

V. **MBASS-SAУ Scenario**

In this section, we consider a general NOMA scheme in which each subcarrier is assigned to \( K \) single-antenna users. We focus on one of the subcarriers and investigate the assignment of spatial dimensions. We consider the case of \( N \leq K \leq 2N \). For such a setting there are at most \( N \) spatial dimensions available. The base station divides the set of users into \( K/2 \) pairs and applies one-step SIC to each pair. Note that the resources in this scenario are the available spatial dimensions.

For ease of notation, assume that \( K \) is an even integer. We model the received signal at the base station as
\[
y = Hx + n,
\]
(27)
where \( y \in \mathbb{C}^N \) is the received vector, \( H \in \mathbb{C}^{N\times K} \) is the channel matrix, \( x \in \mathbb{C}^K \) is the transmit vector of all the users and \( n \sim \mathcal{CN}(0, \sigma_n^2I) \) is additive white Gaussian noise at the base station.
Assume that \( x \) is iid Gaussian with covariance matrix \( \Phi \). The base station first applies a \( K \times N \) detection matrix
\[
    r = Ty. \tag{28}
\]
This pairing is solely done at the base station and the users need not know the pairing strategy.

A. Optimum user pairing

Assume that the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th user are paired for \( m \in \{1, \ldots, K/2\} \). Without loss of generality assume \( \ell_{2m-1} < \ell_{2m} \). Let \( T_m \in \mathbb{C}^{2 \times N} \) consist of the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th row of \( T \). Furthermore, let \( H_m \) consist of the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th column of \( H \) and \( H_m \) is the matrix \( H \) when \( H_m \) is excluded from it. The sum rate of the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th user is calculated as
\[
    R_m \triangleq R_{\ell_{2m-1}} + R_{\ell_{2m}} = \log_2 \det \left( I + T_m H_m \Phi_m H_m^\dagger \Theta_m^{-1} \right), \tag{29}
\]
where \( \Theta_m \) is the covariance matrix of a vector including the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th user,
\[
    \Theta_m \triangleq T_m \left( H_m \Phi_m H_m^\dagger + \sigma_n^2 I \right) T_m^\dagger, \tag{30}
\]
and \( \Phi_m \) is the covariance matrix of the input vector when the \( \ell_{2m-1} \)th and the \( \ell_{2m} \)th entry of it are excluded.

The following theorem is used to obtain the maximum total rate of the users.

**Theorem 5:** Let \( X \in \mathbb{C}^{2 \times K} \) and \( A, B \in \mathbb{C}^{K \times K} \) be two Hermitian matrices. Furthermore, let \( A \) be non-negative with rank 2 and \( B \) be a full rank positive definite matrix. Then,
\[
    \max_X \det \left( I + XAX^\dagger (XBX^\dagger)^{-1} \right) = (1 + \lambda_1)(1 + \lambda_2), \tag{31}
\]
where \( \lambda_1 \) and \( \lambda_2 \) are the nonzero eigenvalues of \( B^{-1/2}AB^{-1/2} \).

**Proof:** From eigenvalue decomposition, we have \( B^{-1/2}AB^{-1/2} = V_1 \Lambda V_1^\dagger \) where \( V_1 \) is a unitary matrix and \( \Lambda \) is a diagonal matrix containing the eigenvalues. By defining \( V \triangleq XB^{1/2} \) and \( U \triangleq VV_1 \), we have
\[
    \max_X \det \left( I + XAX^\dagger (XBX^\dagger)^{-1} \right) = \max_V \det \left( I + V B^{-1/2}AB^{-1/2}V^\dagger (VV^\dagger)^{-1} \right)
    = \max_V \frac{\det \left( V \left( I + B^{-1/2}AB^{-1/2} \right) V^\dagger \right)}{\det (VV^\dagger)}
    = \max_U \frac{\det \left( U \left( I + \Lambda U^\dagger \right) \right)}{\det (UU^\dagger)}.
\]
$B^{-1/2} A B^{-1/2}$ has only two nonzero eigenvalues denoted by $\lambda_1$ and $\lambda_2$. Without loss of generality assume that these two nonzero eigenvalues are the first and the second eigenvalues. Let $U = [U_1 u_2]$ where $U_1 \in \mathbb{C}^{2 \times K-1}$ and $u_2 \in \mathbb{C}^{2 \times 1}$. Using the matrix determinant lemma results in

$$\frac{\det(U (I + \Lambda) U^\dagger)}{\det(UU^\dagger)} = \frac{\det(U_1 (I + \Lambda_1) U_1^\dagger)}{\det(U_1 U_1^\dagger)} \frac{1 + u_2^\dagger (U_1 (I + \Lambda_1) U_1^\dagger)^{-1} u_2}{1 + u_2^\dagger (U_1 U_1^\dagger)^{-1} u_2}. \quad (33)$$

The maximum eigenvalue of $(U_1 (I + \Lambda_1) U_1^\dagger)^{-1}$ is less than or equal to the maximum eigenvalue of $(U_1 U_1^\dagger)^{-1}$ [29]. Therefore, the optimum solution for $u_2$ is the all-zero vector. One can repeat this strategy for the last $K-3$ columns of $U_1$ and show that only the first two columns of the optimum matrix $U$ are nonzero and

$$\max_X \det(I + X A X^\dagger (X B X^\dagger)^{-1}) = (1 + \lambda_1)(1 + \lambda_2). \quad (34)$$

Using Theorem 5, we obtain

$$\max_{T_m} R_m = \log_2(1 + \lambda_{m,1}) + \log_2(1 + \lambda_{m,2}), \quad (35)$$

where $\lambda_{m,1}$ and $\lambda_{m,2}$ are the nonzero eigenvalues of the matrix

$$\Omega \triangleq \left(H_{\setminus m} \Phi_{\setminus m} H_{\setminus m}^\dagger + \sigma_n^2 I\right)^{-1/2} \left(H_m \Phi_m H_m^\dagger \right) \left(H_{\setminus m} \Phi_{\setminus m} H_{\setminus m}^\dagger + \sigma_n^2 I\right)^{-1/2}. \quad (36)$$

Finally, the maximum total rate is derived as

$$\sum_{m=1}^{K/2} R_m = \sum_{m=1}^{K/2} \log_2(1 + \lambda_{m,1}) + \log_2(1 + \lambda_{m,2}). \quad (37)$$

To obtain the optimum pair, we form a $K \times K$ cost matrix whose $(i, j)$th entry denotes the contribution of the $i$th and $j$th users in the total rate which is given in (35) for the $\ell_{2m-1}$th and the $\ell_{2m}$th user. Note that the contribution of the $\ell_{2m-1}$th and the $\ell_{2m}$th user in the total rate should be independent from how the other users, i.e., all the users except the $\ell_{2m-1}$th and the $\ell_{2m}$th user, are grouped. If this condition does not hold, the problem cannot be modeled as a linear assignment problem with a cost matrix. To have such a condition, the eigenvalues $\lambda_{m,1}$ and $\lambda_{m,2}$ should be the same for all pairing schemes in which the $\ell_{2m-1}$th and the $\ell_{2m}$th user are paired. It is straightforward to observe from (36) that this condition is fulfilled.

Having the cost matrix, the problem becomes a linear assignment, and we find the optimum user pairing set using the Hungarian algorithm. The $(\ell_{2m-1}, \ell_{2m})$th and the $(\ell_{2m}, \ell_{2m-1})$th entries
of the cost matrix are equal to $\log_2(1 + \lambda_{m,1}) + \log_2(1 + \lambda_{m,2})$. The Hungarian algorithm gives us $K/2$ pairs with the maximum total rate.

Note that the scheme introduced in this section is not in power domain but in spatial domain. The pairing is done at the base station, and the users need not do anything. At the base station, the users are grouped into pairs and interference cancellation is applied to each pair. The approach introduced here can be used in massive MIMO uplink.

B. Large system analysis

In this subsection, we analyze the NOMA scheme presented in the previous subsection in the large system limit, i.e., $N, K \to \infty$ and constant $\alpha \triangleq K/N$. The goal is to calculate the eigenvalues $\lambda_{m,1}$ and $\lambda_{m,2}$ in the large system limit. We confine the analysis to the case of PPC and leave the case of EP for future works. For PPC, we show that optimum user pairing performs identical to random pairing in the large system limit. However, in the results section, we demonstrate that for finite $K, N$, optimum user pairing performs much better than random pairing. Furthermore, in the results section, using computer simulation we show that the same behavior is observed for EP.

For PPC, we set $P_i = 1/d_i$. Therefore, the path loss effect is canceled out. To analyze such a NOMA system, one can consider a similar NOMA system in which the users have identical power and the channel matrix represents the small scale fading effect. In the following theorem, we present the main results of this subsection.

**Theorem 6:** Let $H$ be an iid matrix whose entries have zero mean and variance of $1/N$. Let $\Phi/\sigma_n^2 = \gamma I$. Then, for $K, N \to \infty$, constant $\alpha = K/N$, PPC, and any pairing method, we have

$$
\lambda_{m,1}, \lambda_{m,2} \xrightarrow{p} \sqrt{\frac{(1 - \alpha)^2\gamma^2}{4} + \frac{(1 + \alpha)\gamma}{2} + \frac{1}{4} - \frac{1}{2} + \frac{(1 - \alpha)\gamma}{2}}.
$$

**Proof:** The proof is given in Appendix B.

Note that Theorem 6 is very general and independent of the distribution of the entries of $H$. In fact, Theorem 6 results from channel hardening, i.e. in the large system limit of PPC there is no channel better than any other. From Theorem 6, it is concluded that for $K, N \to \infty$, we have

$$
\frac{1}{K} \sum_{m=1}^{K/2} R_m \xrightarrow{p} \frac{2}{\alpha} \log_2 \left( \frac{1}{2} + \sqrt{\frac{(1 - \alpha)^2\gamma^2}{4} + \frac{(1 + \alpha)\gamma}{2} + \frac{1}{4} + \frac{(1 - \alpha)\gamma}{2}} \right).
$$

(39)
Theorem 6 shows that in the large system limit, the total rate of PPC converges to a limit independent of the pairing strategy. This limit is equal to the total rate of optimum linear detection and random pairing. However, in finite sizes, in the numerical results section we show that optimum user pairing achieves higher total rate than random pairing. We conjecture that the gain of optimal pairing in NOMA remains in the large system limit if the size of the groups scales with $K$. In such a case, interference cancellation with higher steps is required.

VI. MBASS-MAU SCENARIO

Consider the case that all user terminals and the base station have multi-antenna arrays. Each user has $L$ antennas and the number of antennas at the base station is $N$. It is assumed that all the $K$ users communicate with the base station over a single subcarrier. For such an uplink channel, a NOMA scheme using a signal alignment technique is proposed in [23] in which the users are divided into $K/2$ pairs. Then, the inter-pair interference is mitigated using a combination of beamforming at the user terminals and linear detection at the base station if $K \leq 2N < 4L$ and the channels between the users and the base station are full rank. The base station then applies one-step SIC for every pair. It is shown in [23] that this NOMA system can be decomposed into $K/2$ orthogonal single-antenna NOMA systems.

In this section, we use the same method introduced in the last section to propose a new NOMA uplink scheme in which each user terminal uses a beamforming vector to transmit a single data stream. We show that this method has higher performance than the NOMA scheme with signal alignment proposed in [23]. Furthermore, in contrast to the scheme with signal alignment, in our scheme users only need to know their own channel.

Assume that the $k$th user sends the data stream $u_k$ by applying the beamforming vector $b_k \in \mathbb{C}^{L \times 1}$. The received vector at the base station reads

$$y = \sum_{k=1}^{K} G_k b_k u_k + n,$$

where $G_k$ is the channel matrix between the $k$th user and the base station. In [23], full rank channel matrices and $2L > N$ are assumed to guarantee a user pairing solution. We show that none of these assumptions are required for the method introduced in this section. Note that we have already shown in the previous section that this NOMA method works even for single-antenna users.
It is assumed that the users only know their own channel and the base station knows all the channel matrices. We use a simple beamforming method at the user terminals and leave the optimum beamforming strategy for the future works. The $m$th user sets

$$b_m = v_m,$$  \hspace{1cm} (41)

where $v_m$ is the eigenvector of $G_m^\dagger G_m$ which corresponds to the maximum eigenvalue. Letting

$$\tilde{H} \triangleq [G_1 b_1, \ldots, G_K b_K]$$  \hspace{1cm} (42)

and

$$u \triangleq [u_1, \ldots, u_K]^T,$$  \hspace{1cm} (43)

the received signal at the base station is

$$y = \tilde{H} u + n.$$  \hspace{1cm} (44)

Eq. (44) describes the same channel model as in the case of single-antenna user terminals. Thus, we apply the same detection technique as described in the previous section to pair the users for the case of $K = 2N$. The optimum user pairing set can be obtained using the Hungarian algorithm. Note that in this NOMA scheme, the channel matrices are not required to be full rank. Furthermore, $L$ can be any integer. In the numerical results section, we compare this method against the NOMA scheme with signal alignment.

VII. NUMERICAL AND SIMULATION RESULTS

In this section, some numerical results for the proposed NOMA schemes explained in the previous sections are presented. The cell is assumed to be a disc with radius of $R_d = 100$ and $r_0$ is set to 1. Both the PPC and EQ strategies are considered. We define the average transmit power of the users as

$$\bar{P} \triangleq \mathbb{E} P_i,$$  \hspace{1cm} (45)

and the average path loss of the users as

$$\bar{d} \triangleq \mathbb{E} d_i.$$  \hspace{1cm} (46)
To have a fair comparison, we use the parameter $\bar{\gamma} \triangleq \frac{dP_i}{\sigma_n^2}$ as a measure for SNR at the receiver. Note that $\bar{\gamma}$ is different than $E\gamma_i$. We do not use the measure $E\gamma_i = Ed_iP_i/\sigma_n^2$ for SNR since the power of the users appears with the weight of the path loss coefficients in it.

Besides the results which are obtained from analytical formulas, all the rates are derived using expectation over both the users’ positions in the cell and the small scale fading coefficients. Note that the shadowing effect is neglected here. The rates are calculated using computer simulation with enough samples.

As a reference for comparison, an orthogonal method based on TDMA is used in which the $i$th user is assigned $\zeta_i$ portion of the total time at a given subcarrier. We consider two TDMA cases: 1) when the users at each subcarrier are assigned $\zeta_i = 1/K$ portion of the total time, 2) when $\zeta_i$s are optimized for all the subcarriers.

To measure the fairness in different scenarios, we use the Jain index given by [30]

$$\text{Jain’s fairness index} = \frac{\left( \frac{1}{MK} \sum_{k=1}^{MK} R_k \right)^2}{\frac{1}{MK} \sum_{k=1}^{MK} R_k^2}, \quad (47)$$

which is between 0 and 1. Note that the maximum fairness is obtained when all the rates are equal.

A. SAMS in frequency-selective fading

We present the results for the sub-optimum user pairing method introduced in Section III. Note that the results here are lower bounds for the total rate, since we have used the suboptimum user pairing methods based on the bounds in Lemma 1. Fig. 1 shows the total rate of the users normalized by $M \log_2(1 + 2\bar{\gamma} \log M)$ for $M = 10$ versus $\bar{\gamma}$. It is observed that in both the EP and PPC cases, the proposed pairing methods perform much better than random pairing. The results for the pairing methods based on the first and the second bound in Lemma 1 are almost the same. Thus, in the remaining figures, we only consider the method based on the first bound. It is also observed in Fig. 1 that the NOMA scheme with EP has a better performance than the one with PPC. The reason is that for PPC, the users compensate the path losses, hence the received powers of the users at the base station differ less and accordingly the gain of NOMA reduces. Another observation in Fig. 1 is that for the case of $\bar{\gamma} \to \infty$, the normalized total rate converges to 1 which confirms the predictions of Section III.
Next, we investigate the total rate versus the number of subcarriers. $\bar{\gamma}$ is set to 5 dB. The total rates normalized by $M \log_2 (1 + 2\bar{\gamma} \log M)$ are plotted in Fig. 2. The results in both cases converge to 1 very slowly. Note that the low convergence rate is due to the terms containing $\log_2 \log M$ in the normalized total rate. The results for random pairing are also plotted. As observed, the normalized total rates for random pairing converge to 0 since the total rates in this case scale with $M$. As for the previous figure, the EP method performs better than PPC.

In Fig. 3 we compare the NOMA scheme with two OMA schemes based on TDMA which achieves the same set of rates. $M$ is set to 10. The required total power normalized by the total power of the NOMA scheme is selected as performance measure. This measure shows how much power each method requires to reach the same rate sets as in the NOMA scheme. It is observed that the NOMA schemes are much more power efficient than the two OMA schemes. Note that the NOMA scheme does not need any time management which is very critical in the case of the TDMA method with optimum $\zeta_i$. Furthermore, it is observed that the TDMA methods require more power in the EP case.

Next, we plot the Jain fairness index versus the number of sub-carriers for different scenarios. The results are shown in Fig. 4. As observed, the proposed pairing method performs much
Fig. 2: The total rate normalized by $M \log_2 \log M$ for SAMS in frequency-selective fading and $\bar{\gamma} = 5$ dB versus $M$.

Fig. 3: The normalized total required power of two OMA schemes for SAMS in frequency-selective fading and $M = 10$. The powers at each case are normalized with the total power of the corresponding NOMA scheme.
better in terms of fairness than random pairing method. Furthermore, PPC performs better than EP which is in fact clear because in PPC the path loss is compensated by power control. The main message of this figure is that our proposed user pairing method not only increases the sum rate but also improves fairness in comparison to random pairing.

**B. SAMS in frequency-flat fading**

In this subsection, the results for the case of frequency-flat channels in Section IV are presented. We compare the performance of optimum pairing against random pairing. The results are shown in Fig. 5 for $\bar{\gamma} = 15$ dB and the EP case. The simulation results together with the analytical results $M \to \infty$ are plotted. It is observed that optimum pairing has better performance than random pairing. As a benchmark, the results for full interference cancellation are also plotted which is, in fact, the ergodic capacity per user. In this method, all the users transmit on all subcarriers and the base station uses full interference cancellation. It is easy to show that in such a case for $M \to \infty$, the ergodic capacity per user converges to $\frac{1}{2} \log_2(1 + 2\bar{\gamma}) \approx 3$. From Fig. 5 the rate per user of NOMA with optimum pairing is worse than the ergodic capacity per user. Note, however, that NOMA with one-step interference cancellation has much lower
Full interference cancellation

Optimum, analytical

Optimum, simulation

Random, simulation

Random, analytical

Rate per user (bits/channel use)

Fig. 5: The rate per user of optimum and random user pairing for SAMS in frequency-flat fading, the EP method and $\bar{\gamma} = 15$ dB. The ergodic capacity per user, which is achieved when the full interference cancellation is used, is also plotted.

complexity than full interference cancellation.

Next, we plot the analytical results for $M \to \infty$ versus $\bar{\gamma}$ for optimum and random pairing in both the EP and PPC case. The results are given in Fig. 6. It is observed that optimum pairing performs about 1 dB better than random pairing for both EP and PPC. The EP method performs about 2.5 dB better than the PPC method at around $\bar{\gamma} = 10$ dB. As a benchmark, the ergodic capacities of EP and of PPC are also plotted. The EP method with optimum pairing looses about 4 dB compared to the ergodic capacity. For the PPC method, optimum pairing falls only little behind the ergodic capacity.

We also compare the performance of the NOMA scheme using optimum user pairing against the TDMA schemes in Fig. 7. The number of subcarriers is set to $M = 10$. The normalized required total power to achieve the same set of rates as in the NOMA scheme are plotted versus SNR. The TDMA schemes with $\zeta_i = 0.5$ require much higher power compared to the case of NOMA. This extra power for EP is much larger than for PPC.
Fig. 6: The average rate of the users versus $\bar{\gamma}$ for optimum and random pairing methods, and SAMS in frequency-flat fading.

Fig. 7: The total required power of two OMA schemes normalized by the total power of the NOMA scheme for SAMS in frequency-flat fading and $M = 10$. 
The rate per user for MBASS-SA U. The base station has $N$ antenna and the users have single antenna. $\bar{\gamma}$ is set to 15 dB.

C. MBASS-SA U

The results for optimum user pairing in the case of multi-antenna base stations and one subcarrier are presented in this subsection. The users are assumed to have a single-antenna. The results are plotted in Fig. 8 for $K/N = 2$ and $\bar{\gamma} = 15$ dB. The rate per user for random pairing is also plotted. It is observed that for small and moderate $N$, optimum user pairing outperforms random pairing significantly. For $K, N \to \infty$ and both EP and PPC, the results for both optimum and random pairing converge to the same limit. For PPC, the analytical result for $M \to \infty$, derived based on Theorem 6 for $\alpha = 2$, is also plotted. It is observed that the results for both random and optimum pairing converge to the analytical limit for large $N$.

D. MBASS-MA U

In this subsection, we present the results for the NOMA scheme introduced in Section VI in which the users and the base station are equipped with multiple antennas. The results are plotted for $K = 2N$, $N = 2L + 1$ and $\bar{\gamma} = 10$ dB in Fig. 9 for both EP and PPC. The results of NOMA with user alignment, random user pairing and optimum detection are also plotted for sake of
comparison. We leave optimum user pairing in user alignment based NOMA for future works. It is observed that the method introduced in Section VI with random user pairing performs better than the alignment method with random pairing. Furthermore, optimum user pairing derived by the Hungarian algorithm has much better performance than random user pairing in the proposed NOMA scheme for finite $N$. It can be seen that similar to the case of single-antenna users, random user pairing approaches the performance of optimum user pairing for $N \to \infty$.

VIII. CONCLUSIONS AND FUTURE WORKS

In this paper, NOMA uplink was considered and optimal user pairing was discussed in various settings. For single-antenna base stations and user terminals communicating over $M$ subcarriers with optimal user pairing, it was shown that the total rate scales with $M \log_2 \log M$ if the users have iid small scale fading coefficients. For frequency-flat channels over the subcarriers, the optimal user pairing method was analyzed in the large system limit and shown to be a bit better than random pairing. In both cases, two orthogonal methods were compared against NOMA, and it was shown that the power efficiency of NOMA is much higher than the efficiencies of the orthogonal schemes. Furthermore, a novel NOMA scheme for the case of multi-antenna
base stations was proposed in which the user terminals transmit over a single subcarrier. It was shown that optimum user pairing can be implemented by means of the Hungarian algorithm with polynomial complexity. The results for multi-antenna users were compared against NOMA based on signal alignment. It was shown that a significant performance gain can be achieved.

In this paper, the power allocation was assumed to be fixed for the user terminals. Extending the results of this paper to the case of joint user pairing and power allocation can be interesting future work. One possible approach is to iterate user pairing and power allocation. Another problem for future is the joint assignment of subcarrier and spatial dimension. Furthermore, the methods in this paper were based on perfect channel state information. Investigating the effect of imperfect channel state information and extending the results to the case of user pairing based on large scale fading coefficients for the sake of complexity reduction can be interesting future works.

**APPENDIX A**

**PROOF OF THEOREM**

We first introduce a lower bound and an upper bound for the maximum total rate and then prove the theorem using these two bounds.

**A. The upper bound**

A simple upper bound for the total rate is obtained when the best two users at each subcarrier are selected. Note that this method may not be achievable since some of the users may be among the best users for several subcarriers. Consider the \( m \)th subcarrier and let also \( \kappa_{2m-1} \) and \( \kappa_{2m} \) be the indices of the users with the first and second strongest channels on the \( m \)th subcarrier, i.e.,

\[
\kappa_{2m-1} = \operatorname{argmax}_{i \in \{1, \ldots, 2M\}} \gamma_i |h_{i,m}|^2, \tag{48}
\]

and

\[
\kappa_{2m} = \operatorname{argmax}_{i \in \{1, \ldots, 2M\}, i \neq \kappa_{2m-1}} \gamma_i |h_{i,m}|^2. \tag{49}
\]

Let the support of \( \gamma_i \) be \([c_1, c_2]\) where \( 0 < c_1 < c_2 < \infty \).
Lemma 3: For $M \to \infty$,
\[
c_1 \leq \frac{\gamma_{k2m-1} |h_{k2m-1,m}|^2}{\log(M)}, \quad \frac{\gamma_{k2m} |h_{k2m,m}|^2}{\log(M)} \leq c_2.
\] (50)

Proof:

The proof is straightforward using the Fisher-Tippett-Gnedenko theorem [31] and the fact that $c_1 \leq \gamma_i \leq c_2$ for all $i \in \{1, \ldots, 2M\}$.

Using the result of this lemma, it is concluded that the total rate of this upper bound method is less than or equal to $M \log_2(1 + 2c_2 \log M)$ with probability 1. Since this method gives an upper bound for the optimum total rate, it is concluded that for $M \to \infty$
\[
\max_{(\ell_{m-1}, \ell_m) \in \Pi} \sum_{m=1}^{K} R_m \leq \frac{p}{M \log_2(1 + 2c_2 \log M)} \leq 1.
\] (51)

B. The lower bound

Consider an achievable strategy in which for each subcarrier the first and the second best user among all the users excluding the users already selected for the previous subcarriers, are chosen. This means that we first select the two best users for the first subcarrier and then excluding this two users we select the best two users for the second subcarrier and continue this procedure up to the end. In fact, for the $m$th subcarrier, we select the best two users among $2(M - m + 1)$ users.

Let $\eta_{2m-1}$ and $\eta_{2m}$ denote the indices of the first and the second selected user, respectively, at the $m$th subcarrier. Using a similar method as in the previous subsection, it can be shown that for all $2(M - m + 1)$ where $m = \nu M$ and $\nu \in (0, 1]$ is a non-vanishing constant, both $\frac{\gamma_{\eta_{2m-1}} |h_{\eta_{2m-1},m}|^2}{\log(M-m+1)}$ and $\frac{\gamma_{\eta_{2m}} |h_{\eta_{2m},m}|^2}{\log(M-m+1)}$ are in $[c_1, c_2]$ with probability of 1 when $M \to \infty$. Therefore, the asymptotic behavior of the total rate of the considered method behaves similarly as $\sum_{m=1}^{M} \log_2(1 + 2\hat{c} \log m)$ where $\hat{c}$ is a nonzero positive constant. Next, we present a lemma to investigate the asymptotic behavior of this sum:

Lemma 4: For finite and positive $\hat{c}$, we have
\[
\lim_{M \to \infty} \frac{\sum_{m=1}^{M} \log_2(1 + 2\hat{c} \log m)}{M \log_2(1 + 2\hat{c} \log M)} = 1.
\] (52)

Note that for finite $2(M - m + 1)$, i.e., vanishing $\nu$, the statement may not be true but this does not change the asymptotic behavior.
Proof: The function
\[
  w(x) = \frac{\log_2(1 + a_1x)}{\log_2(1 + a_2x)}
\]
is monotonically increasing if \(a_1 < a_2\) and decreasing if \(a_1 > a_2\) for \(x \geq 0\). For \(a_1 = a_2\), the function becomes constant. Therefore, using L’Hospital’s rule and the Stirling approximation, we have
\[
  \lim_{M \to \infty} \frac{\sum_{m=1}^{M} \log_2(1 + 2\hat{c} \log m)}{M \log_2(1 + 2\hat{c} \log M)} \geq \lim_{M \to \infty} \frac{\sum_{m=1}^{M} \log_2(1 + 2\hat{c} \log M)}{M \log_2(1 + 2\hat{c} \log M)} = \lim_{M \to \infty} \frac{\sum_{m=1}^{M} \log m}{M \log M} = 1.
\]  
Furthermore, since \(m \leq M\) in the summation, the limit is less than or equal to 1. Therefore, the limit is 1.

Therefore, the total rate of the considered pairing method normalized by \(M \log_2 \log M\) converges to 1 in probability.

Finally, based on the results obtained from the above two subsections, it is concluded that
\[
  \max_{(\ell_2m_1, \ell_2m_2) \in \Pi} \sum_{m=1}^{M} R_m \stackrel{p}{\to} 1
\]
for \(M \to \infty\).

APPENDIX B

PROOF OF THEOREM 6

For such a setting, we have
\[
  \Omega = \left( H_{\setminus m} H_{\setminus m}^\dagger + I/\gamma \right)^{-1/2} H_m H_m^\dagger \left( H_{\setminus m} H_{\setminus m}^\dagger + I/\gamma \right)^{-1/2}.
\]  
The two nonzero eigenvalues of \(\Omega\) are equal to the eigenvalues of
\[
  \tilde{\Omega} = H_m^\dagger \left( H_{\setminus m} H_{\setminus m}^\dagger + I/\gamma \right)^{-1} H_m,
\]
which is a \(2 \times 2\) matrix. For \(N, K \to \infty\), the off-diagonal entries of \(\tilde{\Omega}\) converge to zero due to the fact that the entries of \(H_m\) and \(H_{\setminus m}\) are iid and zero mean. Furthermore, the two diagonal entries of \(\tilde{\Omega}\) fulfill [32, Lemma 2.29]
\[
  \tilde{\Omega}_{1,1} \to \tilde{\Omega}_{2,2} \to \tilde{G} \left( -\frac{1}{\gamma} \right),
\]
where
\[
G(w) = \sqrt{\frac{(1 - \alpha)^2}{4w^2} - \frac{1 + \alpha}{2w} + \frac{1}{4} - \frac{1}{2} - \frac{1 - \alpha}{2w}}
\]
is the Stieltjes transform of the eigenvalue distribution of $H_m H_m^\dagger$. Therefore, the two eigenvalues of $\hat{\Theta}$ are obtained as
\[
\lambda_{m,1} \rightarrow \lambda_{m,2} \rightarrow \sqrt{\frac{(1 - \alpha)^2 \gamma^2}{4} + \frac{(1 + \alpha) \gamma}{2} + \frac{1}{4} - \frac{1}{2} + \frac{(1 - \alpha) \gamma}{2}},
\]
for $K, N \rightarrow \infty$.

REFERENCES

[1] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in IEEE 77th Vehicular Technology Conference (VTC Spring), 2013, pp. 1–5.

[2] 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 Processing Letters, vol. 21, no. 12, pp. 1501–1505, 2014.

[3] Z. Ding, P. Fan, and V. Poor, “Impact of user pairing on 5G non-orthogonal multiple access downlink transmissions,” IEEE Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[4] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, “Application of non-orthogonal multiple access in LTE and 5G networks,” IEEE Communications Magazine, vol. 55, no. 2, pp. 185–191, 2017.

[5] Z. Chen, Z. Ding, X. Dai, and R. Zhang, “A mathematical proof of the superiority of NOMA compared to conventional OMA,” arXiv preprint arXiv:1612.01069, 2016.

[6] H. Tabassum, M. S. Ali, E. Hossain, M. Hossain, D. I. Kim et al., “Non-orthogonal multiple access (NOMA) in cellular uplink and downlink: Challenges and enabling techniques,” arXiv preprint arXiv:1608.05783, 2016.

[7] S. Timotheou and I. Krikidis, “Fairness for non-orthogonal multiple access in 5G systems,” IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1647–1651, 2015.

[8] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding, “A survey of downlink non-orthogonal multiple access for 5G wireless communication networks,” arXiv preprint arXiv:1609.01856, 2016.

[9] 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), 2013 IEEE. IEEE, 2013, pp. 66–70.

[10] H. Nikopour and H. Baligh, “Sparse code multiple access,” in Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on. IEEE, 2013, pp. 332–336.

[11] H. Nikopour, E. Yi, A. Bayesteh, K. Au, M. Hawryluck, H. Baligh, and J. Ma, “SCMA for downlink multiple access of 5G wireless networks,” in Global Communications Conference (GLOBECOM), 2014 IEEE. IEEE, 2014, pp. 3940–3945.

[12] D. Tse and P. Viswanath, Fundamentals of Wireless Communications. New York: Cambridge University Press, 2005.

[13] G. Caire, R. R. Müller, and R. Knopp, “Hard fairness versus proportional fairness in wireless communications: The single-cell case,” IEEE Transactions on Information Theory, vol. 53, no. 4, pp. 1366–1385, 2007.

[14] K. Higuchi and A. Benjebbour, “Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access,” IEICE Transactions on Communications, vol. 98, no. 3, pp. 403–414, 2015.
[15] Z. Ding, M. Peng, and H. V. Poor, “Cooperative non-orthogonal multiple access in 5G systems,” *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, 2015.

[16] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for MC-NOMA systems,” *arXiv preprint arXiv:1603.08132*, 2016.

[17] J. He, Z. Tang, and Z. Che, “Fast and efficient user pairing and power allocation algorithm for non-orthogonal multiple access in cellular networks,” *Electronics Letters*, vol. 52, no. 25, pp. 2065–2067, 2016.

[18] R. R. Müller, *Power and Bandwidth Efficiency of Multiuser Systems with Random Spreading*. Aachen, Germany: Shaker–Verlag, 1999.

[19] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, “Uplink non-orthogonal multiple access for 5G wireless networks,” in *11th International Symposium on Wireless Communications Systems (ISWCS)*. IEEE, 2014.

[20] M. Al-Imari, P. Xiao, and M. A. Imran, “Receiver and resource allocation optimization for uplink NOMA in 5G wireless networks,” in *Wireless Communication Systems (ISWCS), 2015 International Symposium on*. IEEE, 2015, pp. 151–155.

[21] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, “On the ergodic capacity of MIMO NOMA systems,” *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 405–408, 2015.

[22] 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, 2016.

[23] 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, pp. 4438–4454, June 2016.

[24] Z. Ding and H. V. Poor, “Design of massive-MIMO-NOMA with limited feedback,” *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 629–633, 2016.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[26] H. W. Kuhn, “The hungarian method for the assignment problem,” *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[27] E. L. Lawler, *Combinatorial optimization: networks and matroids*. Courier Corporation, 2001.

[28] I. Makohon, M. Cetin, D. T. Nguyen, and M. Ng, “Hungarian optimum assignment algorithm with Java computer animation,” in *SoutheastCon*, 2016. IEEE, 2016, pp. 1–7.

[29] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.

[30] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984, vol. 38.

[31] H. A. David and H. N. Nagaraja, *Order statistics*. Wiley Online Library, 1981.

[32] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers Inc, 2004, vol. 1.

[33] R. R. Müller, G. Alfano, B. M. Zaidel, and R. de Miguel, “Applications of large random matrices in communications engineering,” *arXiv preprint arXiv:1310.5479*, 2013.