Joint user pairing and power allocation scheme based on transmission mode switching between NOMA based maximum ratio transmission and MMSE beamforming in downlink MISO systems

Zhixin Zhao (✉ Zhixin_Z@lingnan.edu.cn)  
Lingnan Normal University  https://orcid.org/0000-0002-8915-4924

Dong Wang  
Lingnan Normal University

Hongwei Zhang  
Lingnan Normal University

Haitao Sang  
Lingnan Normal University

Research Article

Keywords: MISO, NOMA, MMSE-BF, rate region, duality, weighted sum rate

Posted Date: February 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-58837/v2

License: ☒  This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Mobile Information Systems

Joint user pairing and power allocation scheme based on transmission mode switching between NOMA based maximum ratio transmission and MMSE beamforming in downlink MISO systems

Zhixin Zhao¹, Dong Wang², Hongwei Zhang¹ and Haitao Sang¹

¹School of Information Engineering, Lingnan Normal University, 524048, China.
²School of Life Science & Technology, Lingnan Normal University, 524048, China.

Correspondence should be addressed to Zhixin Zhao; zhaozhixin0830@gmail.com

Abstract

This paper investigates the design of the joint user pairing and power allocation scheme with transmission mode switching (TMS) in downlink multiple-input-single-output (MISO) systems. Firstly, the closed-form expressions of the boundary of achievable rate region of two candidate transmission modes, i.e., non-orthogonal multiple access based maximum ratio transmission (NOMA-MRT) and minimum mean square error beamforming (MMSE-BF), are obtained. By obtaining the outer boundary of the union of the achievable rate regions of the two transmission modes, an adaptive switching method is developed to achieve a larger rate region. Secondly, based on the idea that the solution to the problem of weighted sum rate (WSR) optimization must be on the boundary of the achievable rate region, the optimal solutions to the problem of WSR optimization for NOMA-MRT and MMSE-BF are obtained for the two-user case, respectively. Subsequently, by exploiting the optimal solutions aforementioned for two transmission modes and the high efficiency of TMS, a suboptimal User pairing and Power Allocation algorithm (JUPA) is proposed to further improve sum-rate performance for the multi-user case. Compared with the Exhaustive Search based user Pairing and Power Allocation algorithm (ES-PPA), the proposed JUPA can enjoy a much lower computational complexity and only suffer a slight sum-rate performance loss, whereas outperforms other traditional schemes. Finally, numerical results are provided to validate the analyses and the proposed algorithms.

1. Introduction

With the continuous emergence of new application scenarios, one of challenges faced by the future wireless communication systems is how to provide higher-speed downlink transmission, restricted to the scarce spectrum resources. The traditional downlink transmission schemes used in mobile communication systems are based on orthogonal multiple access (OMA) technology, e.g., frequency division multiple access (FDMA) for the first generation (1G), time division multiple access (TDMA) for 2G, code division multiple access (CDMA) for 3G, and orthogonal frequency division multiple access (OFDMA) [1, 2] for 4G. These conventional multiple access schemes can mitigate or avoid the inter-user interference by allocating orthogonal resources (frequency/time/code) to different users, which result in insufficient use of spectrum resources.
Recently, Non-orthogonal multiple access (NOMA) has been considered as a promising multiple access scheme for 5th Generation (5G) wireless communication systems, owning to its higher spectrum efficiency compared with the conventional orthogonal multiple access schemes [3-5]. Note that Third Generation Partnership Project (3GPP) has considered NOMA as a study-item for 5G new radio (NR) in Release 15 and decided to leave it for possible use in Beyond 5G (B5G) [6]. NOMA enables multiple users to share the same time-frequency resource block on the same spatial layer. NOMA is achieved by the combination of superposition encoding and successive interference cancellation (SIC), which is a method to reach the boundary of the capacity region of degraded broadcast channel [7].

In order to further enhance system performance, beamforming (BF) was combined with NOMA in multiple-input-single-output (MISO) downlink [8-14]. In [8], the sum rate maximization problem with BF vector being the optimization variable was studied and an one-dimensional iterative algorithm was proposed. However, for each iteration, a second-order cone program convex problem needs to be solve, which results in very high computational complexity. Moreover, as the signals of all users are superposed on one resource block, the algorithm may suffer large process delay and error propagation of SIC for the system with a large number of users. In [9], the sum rate optimization problem with minimum user rate constrain was investigated, and therefore a low complexity BF scheme and a user clustering scheme were proposed. Since the presented BF scheme and user clustering scheme were designed separately, some sum-rate performance loss is suffered. In [10], the problem for maximization of the number of users with an ergodic user rate constrain was considered. A power allocation scheme to satisfy ergodic user rate constrains was proposed and then a user admission algorithm that achieves the maximum number of users was developed to guarantee the minimum user rate requirements. However, as the MRT beamforming is adopted for downlink transmission, the optimization of BF vector is not considered. By the duality between the multiple access channel (MAC) and broadcast channel (BC), the duality scheme for sum rate optimization was developed in [11], which also suffers rather high computational complexity because of needing to solve a quadratic constrained quadratic programs convex problem. Furthermore, since the duality scheme is a quasi-degraded solution to the problem of sum rate optimization actually, it is only feasible in the case, where the channel state information (CSI) of the scheduled users meet the quasi-degrade property. As a result, the application of the duality scheme in practical systems may be heavily restricted. In [12], the robust BF design problem to optimize the worst-case achievable sum rate constrained by the total transmit power was studied. In [13], the optimal BF design problem which minimizes the total transmission power subject to a pair of target interference levels constrains, was investigated for two-user MISO-NOMA downlink. Based on the results obtained by [13], [14] further proved that the minimum transmit power of the NOMA transmission scheme was equal to that of dirty-paper coding in two-user case, under the condition of the broadcast channel being quasi-degraded. Furthermore, a Hybrid NOMA (H-NOMA) precoding algorithm with low-complexity, is proposed by combining NOMA with zero-forcing beamforming (ZFBF). The summary for comparison of this paper with the existing works [8-14], is shown in Table 1.

| Research Works | [8] | [9] | [10] | [11] | [12] | [13] | [14] | This paper |
|----------------|-----|-----|------|------|------|------|------|----------|
| Main objectives | MU MISO | MU MISO | MU MISO | TU MIMO | MU MISO | TU MIMO | MU MISO | MU MISO |
| Optimization variable | Sum rate | Sum rate | Number of users | Sum rate | WC sum rate | Transmit power | Transmit power | Sum rate |
| Joint optimization | BV | BV+CM | Power | Precoding matrix | BV+power | BV | BV+CM | BV+ UP +CP |
| User-number in a cluster | – | No | – | – | No | – | Yes | Yes |
| MAC-BC duality | Not use | Not use | Not use | Use | Not use | Not use | Not use | Use |
| Each BF vector | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
Dynamic user clustering and user pairing are effective methods to achieve the viable benefits of NOMA in providing high spectral efficiency [15-17]. In [15], the impact of user pairing on the performance of the two-user NOMA systems, is characterized. It was shown that the performance gain of NOMA over conventional multiple access can be further enlarged by electing users whose channel conditions are more distinctive. In [16], the problem of jointly optimizing user association and power control to maximize the overall spectral efficiency, was investigated for NOMA-Based Full-Duplex MIOS Systems. To be spectrally-efficient, the tensor model was introduced to optimize uplink users’ decoding order and downlink users’ clustering. In [17], a novel hybrid user pairing beamforming (HUP) scheme was proposed to maximize the achievable spectral and energy efficiencies for Multiuser MISO-NOMA downlink systems with simultaneous wireless information and power transfer (SWIPT). In the proposed HUP scheme, two information users with distinct channel conditions are paired while energy users located nearby the BS perform energy harvesting.

Due to the equivalence between the maximization of the weighted sum rate (WSR) and the acquisition of the maximum WSR point on achievable rate region of the MISO downlink, the problems of characterization of the achievable rate region for MISO broadcast channel [18,19] and for MISO interference channel [20-22], were investigated, respectively. Specifically, in [18], the set of BF vectors which achieve points on the boundary of achievable rate region of two-user MISO broadcast channel, are characterized by a single real valued parameter per user. In [19], the design of adaptive transmission mode switching to derive the larger rate region for the two-user MISO broadcast channel is investigated. An explicit characterization of the boundary of achievable rate region for multiuser MISO interference channel was obtained in [20]. A general framework for finding the maximum sum-rate operating points on the boundary of the achievable rate region for the two-user MISO interference channel, was proposed in [21]. In [22], the achievable rate region of two-user MISO interference channel for single user detection, was obtained.

Although the aforementioned schemes can provide efficient solutions with several advantages, they bring few insights about their optimality, compared to the achievable rate regions for downlink MISO systems with transmission mode switching. In this paper, we compare the achievable rate regions obtained by NOMA-MRT and MMSE-BF. Then, we find that when the channel vectors of the two users are greatly correlated, the achievable rate region of NOMA-MRT include that of MMSE-BF completely. However, when the two channels are almost orthogonal, the opposite is true. Consequently, we focus on the design of adaptive transmission mode switching (TMS) between NOMA-MRT and MMSE-BF, to maximize the sum rate for multi-user MISO downlink, by utilizing the idea of user pairing.

The main contributions for this paper is listed as follows:

1) By combining MRT beamforming and NOMA, a novel transmission scheme referred to as NOMA-MRT for MISO downlink is proposed and the corresponding closed-form expression of the boundary of the achievable rate region is achieved. Moreover, building on the duality of MAC and BC, the closed-form expression for the boundary of the achievable rate region of MISO broadcast channel with MMSE-BF adopted at BS, is given. Subsequently, by obtaining the outer boundary of the union of the achievable rate regions of the two transmission schemes, an adaptive transmission mode switching method is developed to achieve a larger rate region for the two-user case.

2) Based on the idea that the solution to the problem of WSR optimization must be on the corresponding boundary of the achievable rate region and the closed-form expression of the rate region boundary in any channel condition, the global maximum value of the transformed-WSR (the unary function) can be found by selecting the maximum among the function values at the end of the interval and the local maximum values of
WSR. Consequently, two optimal power allocation algorithms, i.e., Power Allocation algorithm for NOMA-MRT mode (NOMA-MRT-PA) and Power Allocation algorithm for MMSE-BF mode (MMSE-BF-PA), are proposed by focusing on the case with two users.

3) Building on the NOMA-MRT-PA and MMSE-BF-PA, a novel Joint User pairing and Power Allocation algorithm (JUPA) is proposed for the multi-user case. Subsequently, by the combination of JUPA and the idea of TMS between NOMA-MRT and MMSE-BF, a practical transmission method is developed. By making use of the closed-form expression of the optimal solution by NOMA-MRT-PA (MMSE-BF-PA), JUPA can be performed with a low computational complexity, while the Exhaustive Search based user Pairing and Power Allocation algorithm (ES-PPA) requires extremely high complexity, compared with which JUPA only suffers a slight performance loss.

The remainder of this paper is organized as follows. Section 2 briefly describes the system model and introduces the NOMA-based beamforming scheme. In Section 3, the closed-form expressions of the achievable rate region boundary of MISO broadcast channel by using NOMA-MRT and MMSE-BF is obtained. Consequently, an adaptive switching method is proposed in Section 4. In Section 5, two optimal power allocation algorithms for NOMA-MRT mode and MMSE-BF mode are presented, by which a Joint User pairing and Power Allocation algorithm is also developed, respectively. The numerical results are illustrated in Section 6, and finally conclusions are drawn in Section 7.

Before proceeding, we introduce the following notation. Throughout the paper, we denote column vectors $\mathbf{x}$ and matrices $\mathbf{X}$ by bold lower-case and upper-case letters, respectively. $(\cdot)^H$ represents the complex conjugate transpose of a vector or matrix. The absolute value of a scalar is denoted by $|\cdot|$ and the norm of a vector is by $\|\cdot\|$. $\mathbf{x} \in \mathbb{C}^{M\times1}$ means that $\mathbf{x}$ is an $M \times 1$ complex vector. $\mathcal{CN}(\mu,\sigma^2)$ denotes a complex Gaussian random variable with mean $\mu$ and variance $\sigma^2$. $(\mathbf{x},\mathbf{y})$ and $\angle(\mathbf{x},\mathbf{y})$ denotes the inner product and the angle of two complex vectors $\mathbf{x}$ and $\mathbf{y}$, respectively.

2. Problem Description

2.1 System Model

We consider a downlink communication system with one $M$-antenna base station (BS) and $2K$ (assumed to be an even number) single-antenna users, from which $K$ user-pairs are selected and the two users in a user-pair share the same spectrum. The BS serves one user-pair in each time-frequency resource block and the data transmission for the user-pair can adaptively switch between two transmission modes, i.e., NOMA-MRT and MMSE-BF, according to the channel state, as shown in Figure 1. When the channel vectors of the two users in a user-pair are greatly correlated, NOMA-MRT mode is usually selected to achieve larger sum rate in the system. However, when the two channels are almost orthogonal, MMSE-BF mode is selected normally (See section 6.1 for details). In NOMA-MRT mode, MRT beamforming is performed at the BS. $2K$ users are assumed to be uniformly located within a cell with a radius of $R$, and the BS is deployed at the center of the cell. Here, we consider user pairing, i.e., selecting two users to form a group, in the system model for the following reasons. In the NOMA downlink, users cancel the co-channel interference by performing SIC. However, with the number of users in a group growing, the processing complexity and delay at users dramatically increase [23]. As a result, we consider only grouping two users and adopting user pairing scheme to decrease processing complexity and delay in the NOMA downlink. The observation at user $i$ in the user-pair $j$ is given by
\[ y_{i,j} = h_{i,j}^H x + z_{i,j}, \quad i = m, n; j = 1, 2, \ldots K, \]

where \( h_{i,j} \sim \mathcal{CN}(0, \sigma_{i,j}^2 I_m) \) is the channel vector from BS to user \( i \) in the user-pair \( j \), \( z_{i,j} \sim \mathcal{CN}(0, N_0) \) is the additive white Gaussian noise (AWGN) at user \( i \) in the user-pair \( j \), and \( N_0 \) is the corresponding noise power. For the channel vector \( h_{i,j} \), the variance of the channel from BS to user \( i \) in the user-pair \( j \) is modeled as \([14], [24]\)

\[ \sigma_{i,j}^2 = \begin{cases} d_{i,j}^{-\alpha}, & \text{if } d_{i,j} > d_0, \\ d_0^{-\alpha}, & \text{otherwise,} \end{cases} \]

where \( d_{i,j} \) denotes the distance between BS and user \( i \) in the user-pair \( j \), \( \alpha \) denotes the path loss exponent and the parameter \( d_0 \) avoids the singularity when \( d_{i,j} \) is small. Furthermore, \( x_j = \sqrt{p_m}w_{m,j} s_j + \sqrt{p_n}w_{n,j} n_j \) is the signal transmitted by the BS for the user-pair \( j \), where \( s_j \) and \( w_{i,j} \) are the scalar signal and normalized BF vector for user \( i \) in the user-pair \( j \), respectively. \( p_{i,j} \) is the transmit power allocated to user \( i \) in the user-pair \( j \). Assume that the BS can obtain the perfect CSI.

For convenience, we omit the subscript \( j \) of the variables mentioned above in the following. In this paper, we attempt to maximize the sum rate of users by jointly optimizing the pairing relationship matrix \( U \) as well as power allocation vector \( p_{m,n} \) and control parameter \( t_{m,n} \) for TMS of each user-pair. Mathematically, the optimization problem can be formulated as

---

(a) NOMA-MRT mode

(b) MMSE-BF mode

Figure 1 System model with transmission mode switching between NOMA-MRT and MMSE-BF
(P1): \[ \max_{p_{m,n} \in [0,1], \ T_{m,n} \in \{0,1\}, \ 1 \leq m,n \leq 2K, \ u_{m,n} \in \{0,1\}, \ 1 \leq m,n \leq 2K, \ u_{m,n} = u_{n,m}, \ 1 \leq m,n \leq 2K, \ u_{m,m} = 0, \ 1 \leq m \leq 2K, \sum_{m=1}^{2K} u_{m,n} = 1, \ 1 \leq n \leq 2K, \sum_{n=1}^{2K} u_{m,n} = 1, \ 1 \leq m \leq 2K, 0 \leq u_{m,n}(p_m + p_n) \leq P, 1 \leq m,n \leq 2K, \|w_m\| = \|w_n\| = 1, 1 \leq m,n \leq 2K, \]

where \( R_i \) is the date rate of user \( i \) in the user-pair \( \{m,n\}, i = m,n \). \( U \) is the user pairing relationship matrix with 2K-dimension, in which the \( m \)-th row and \( n \)-th column element is denoted by \( u_{m,n} \). \( p_{m,n} = [p_m, p_n] \) and \( S_{m,n} \) are power allocation vector and control parameter for transmission mode switching of the user-pair \( \{m,n\} \). As shown in (2b), \( S_{m,n} = 1 \) denote that the transmission mode NOMA-MRT is select for the user pair \( \{m,n\} \) at BS. Otherwise, MMSE-BF is performed for the user pair \( \{m,n\} \). (2c) \( u_{m,n} = 1 \) represents that user \( m \) and \( n \) are paired together and constitute a user-pair. Otherwise, user \( m \) and \( n \) are not paired together. (2d) and (2e) imply that \( U \) is a symmetric matrix, and the diagonal elements are zero as a user can’t be paired with itself. (2f) and (2g) indicate that a user can be paired with only one user. Furthermore, (2h) represents the total power allocated to a user-pair \( \{m,n\} \), which is upper bounded to \( P \). (2i) denotes that the BF vectors are normalized in this paper.

It is hard to achieve the optimal solution to the problem (P1) due to binary constraints of \( u_{m,n} \) and the nonconvex property of the achievable rate of user \( i \), which is the data rate of user \( i \) with the equality holding in (3) or (5). Therefore, for the treatability of the problem (P1), we select the BF scheme (i.e., control parameter for TMS) with the larger sum rate for a user-pair, among MMSE-BF and NOMA-based BF, instead of jointly optimizing BF vector for all users, in the process of jointly optimizing the pairing relationship matrix \( U \), power allocation vector \( p_{m,n} \) and control parameter \( T_{m,n} \) of each user-pair.

### 2.2 Achievable rate region of the existing BF Schemes

For conventional linear BF scheme, such as ZFBF, MRT and MMSE-BF, given a fixed normalized BF vector \( w_i \), the achievable rate region of the MISO broadcast channel (BC) is given by the set of rate tuples \( (R_m, R_n) \) satisfying

\[ R_i \leq \log(1 + \gamma_i^B), \]

where

\[ \gamma_i^B = \frac{p_i |h_i^T w_i|^2}{N_0 + p_{\phi(i)} |h_i^T w_{\phi(i)}|^2}, \]

is the signal-to-interference-plus-noise ratio (SINR) of user \( i \) with \( p_m + p_n \leq P, i = m,n \). The mapping function \( \phi(i) \) is defined by
\[ \varphi(i) = \begin{cases} n, & i = m, \\ m, & i = n, \end{cases} \]

and \( P \) is the total transmit power for each user-pair. When the equality holds in (3), the rate tuple \((R_m, R_n)\) achieves on the boundary of the achievable rate region with \( p_m + p_n = P \).

### 2.3 Achievable rate region of NOMA-based BF Scheme

For two-user MISO downlink with the NOMA-based linear BF scheme adopted at the BS, SIC is implemented at the user with strong channel conditions. Therefore, for a fixed normalized BF vector \( \mathbf{w}_i \), when SIC is carried out at user \( n \) with strong channel conditions, the achievable rate region of the MISO broadcast channel by using NOMA-based linear BF at the BS, can be formulated by the set of rate tuples satisfying

\[
\begin{align*}
R_n &\leq \log \left( 1 + \frac{p_m |\mathbf{h}_m^H \mathbf{w}_m|^2}{N_0} \right), \\
R_m &\leq \min \{ \log(1 + \gamma_{m,n}^B), \log(1 + \gamma_m^B) \},
\end{align*}
\]

where \( \log(1 + \gamma_{m,n}^B) \) denotes the achievable rate for user \( n \) to detect user \( m \)'s message and \( \gamma_{m,n}^B \) is the corresponding SINR, i.e.,

\[
\gamma_{m,n}^B = \frac{p_m |\mathbf{h}_m^H \mathbf{w}_m|^2}{N_0 + p_n |\mathbf{h}_n^H \mathbf{w}_n|^2}.
\]

In this paper, the angle between \( \mathbf{h}_m \) and \( \mathbf{h}_n \) is denoted by \( \alpha \), where

\[
\cos^2(\alpha) = \frac{|\mathbf{h}_m^H \mathbf{h}_m|^2}{\| \mathbf{h}_m\|^2 \cdot \| \mathbf{h}_n\|^2}.
\]

Without loss of generality, we set \( \alpha \in [0, \pi) \).

### 2.4 The equivalence between maximizing WSR and enlarging achievable rate region

For a given weight vector, the solution of optimization problem of maximizing WSR must be on the boundary of achievable rate region of the two-user MISO broadcast channel [19]. Thus, maximizing the WSR in MISO downlink systems is equivalent to enlarging the corresponding achievable rate region as much as possible.

In this paper, from the perspective of achievable rate region, we investigate the design of the suboptimal algorithm for the solution to the problem (P1) based on transmission mode switching between NOMA-based BF and MMSE-BF.

### 3. Analysis on Rate Regions of NOMA-MRT and MMSE-BF

#### 3.1 Achievable Rate Region of NOMA-MRT

Geometrically, the BF vector \( \mathbf{w}_i \) of user \( i \) with MRT is aligned with the spatial direction of \( \mathbf{h}_i \) to maximize the length of the projection of \( \mathbf{w}_i \) onto \( \mathbf{h}_i \). Therefore, the maximum of the signal-to-noise ratio for user \( i \) can be achieved by using MRT. The normalized BF vector of user \( i \) by using MRT is given by

\[
\mathbf{w}_i = \frac{\mathbf{h}_i}{\| \mathbf{h}_i \|}.
\]
NOMA-MRT is the combination of NOMA and MRT, i.e., performing MRT beamforming at the transmitter and executing SIC at the receiver with strong channel conditions for two-receiver case.

**Lemma 1** When $\|h_m\| < \|h_n\|$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as

**Case 1) $\theta > 0$**

$$
\begin{align*}
    r_n &= \log \left( 1 + \frac{\bar{p}_n \|h_n\|^2}{1 + \bar{p}_n \|h_n\|^2 (1 - \theta)} \right) \zeta_{n,m} \leq 0 \\
    r_m &= \log \left( 1 + \frac{\bar{p}_m \|h_m\|^2}{1 + \bar{p}_n \|h_n\|^2 (1 - \theta)} \right) \zeta_{n,m} \geq 0
\end{align*}
$$

(7)

where

$$
\zeta_{n,m} = \frac{\|h_n\|^2 (1 - \theta) - \|h_m\|^2}{\|h_m\|^2 \|h_n\|^2 (2\theta - \theta^2)}.
$$

(8)

$\theta = \sin^2 \alpha$, $\rho = P / N_0$, $\bar{p}_m = \rho - \bar{p}_n$ and $\bar{p}_i = p_i / N_0, i = m, n$.

**Case 2) $\theta = 0$**

$$
\begin{align*}
    r_n &= \log \left( 1 + \bar{p}_n \|h_n\|^2 \right) \\
    r_m &= \log \left( 1 + \frac{\bar{p}_m \|h_m\|^2}{1 + \bar{p}_n \|h_n\|^2} \right)
\end{align*}
$$

(9)

Proof See the Appendix.

**Lemma 2** When $\|h_m\| = \|h_n\| = l$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as

$$
\begin{align*}
    r_n &= \log \left( 1 + \frac{p_n l^2}{N_0} \right) \\
    r_m &= \log \left( 1 + \frac{p_m l^2}{N_0 + p_n l^2} \right)
\end{align*}
$$

(10)

Proof See the Appendix.

### 3.2 Achievable Rate Region of MMSE-BF

The MMSE-BF can optimally trade off fighting interference to other users and the background Gaussian noise, i.e., the MMSE-BF can maximize the output SINR for any value of signal to noise ratio (SNR) [25]. Such a beamformer looks like the Zero-Forcing beamformer when the inter-user interference is large and like the MRT
beamformer when the interference is small. The geometric description of normalized BF vectors for ZFBF, MRT and MMSE-BF is shown in Figure 2.

![Figure 2: The geometric description of BF vectors for ZFBF, MRT and MMSE-BF](image)

By MAC-BC duality theory [25], the MMSE beamformer in BC, is exactly the MMSE receive filters in dual MAC. Therefore, the sets of achievable SINRs are the same in both cases with the same total transmit power constrain. Consequently, we have the following Lemma.

**Lemma 3** By the duality of MAC and BC, the achievable rate region boundary of BC by using MMSE-BF can be written in terms of that of the dual MAC with MMSE receive filter. The set of rate tuple $(\bar{r}_m, \bar{r}_n)$ on the achievable rate region boundary of BC satisfies

$$\bar{r}_i = \log(1 + y_i^H), \quad i = m, n,$$

where

$$y_i^H = \frac{\rho_i \| h_i \|^2 (1 + \rho_i \| h_{\varphi(i)} \|^2 \theta)^2}{\gamma_i + \rho_i \| h_{\varphi(i)} \|^2 (1 - \theta)}.$$  

$s_i = \rho_i^2 \| h_{\varphi(i)} \|^4 \theta + 2 \rho_i \| h_{\varphi(i)} \|^2 \theta + 1$. $\rho_i = q_i / N_0$ and $q_i$ is the transmit power of user $i$ in the dual MAC, satisfying $q_i + q_{\varphi(i)} = P$. 

**Proof** According to the duality between MAC and BC [25], the normalized BF vectors of user $i$ with MMSE-BF in MISO broadcast channel is given by

$$w_i^{\text{MMSE}} = \frac{(N_0 I + q_{\varphi(i)} h_{\varphi(i)} h_{\varphi(i)}^H)^{-1} h_i}{\| (N_0 I + q_{\varphi(i)} h_{\varphi(i)} h_{\varphi(i)}^H)^{-1} h_i \|}.$$  

By the matrix inversion Lemma [26], Eq. (13) can be written as
The SINR $\gamma_i^M$ of user $i$ in dual MAC with MMSE receiver filter have the following form

$$\gamma_i^M = \frac{q_i \|h_i^M w_i^{MMSE}\|^2}{N_0 + q_{\phi(i)} \|h_{\phi(i)}^H w_i^{MMSE}\|^2}.$$  \hspace{1cm} (15)

According to the definition of inner product, Eq. (15) can be rewritten as

$$\gamma_i^M = \begin{cases} 
\frac{q_i \|h_i\|^2 \cos^2 \beta_i}{N_0 + q_{\phi(i)} \|h_{\phi(i)}\|^2 \cos^2 (\alpha + \beta_i)} & \alpha \in [0, \pi/2) \\
\frac{q_i \|h_i\|^2 \cos^2 \beta_i}{N_0 + q_{\phi(i)} \|h_{\phi(i)}\|^2 \cos^2 (\alpha - \beta_i)} & \alpha \in [\pi/2, \pi) 
\end{cases}$$  \hspace{1cm} (16)

where $\beta_i$ represent the angle between $w_i^{MMSE}$ and $h_i$, $\alpha + \beta_i$ is the angle between $w_i^{MMSE}$ and $h_{\phi(i)}$ when $\alpha \in [0, \pi/2)$ and $\alpha - \beta_i$ is that when $\alpha \in [\pi/2, \pi)$, $i = n, m$, as illustrated in Figure 2.

The square of the cosine value of $\beta_i$ can be expressed as

$$\cos^2 \beta_i = \frac{\| (w_i^{MMSE}, h_i) \|^2}{\| h_i \|^2}$$  \hspace{1cm} (18)

By substituting (14) into (18), we can derive

$$\cos^2 \beta_i = \frac{\| h_i \|^2 (1 + \rho_{\phi(i)} \| h_{\phi(i)} \|^2)^2}{\| s_i \|^2 (1 + \rho_{\phi(i)} \| h_{\phi(i)} \|^2)^2}.$$  \hspace{1cm} (19)

where

$$s_i = (1 - \frac{\rho_{\phi(i)} h_{\phi(i)}^H h_i}{1 + \rho_{\phi(i)} \| h_{\phi(i)} \|^2}) h_i.$$  \hspace{1cm} (20)

In according to the definition of inner product, $s_i$ can be written as

$$s_i = \frac{h_i - \frac{\rho_{\phi(i)} \| h_{\phi(i)} \|^2}{1 + \rho_{\phi(i)} \| h_{\phi(i)} \|^2} \| h_{\phi(i)} \|^2}{1 + \rho_{\phi(i)} \| h_{\phi(i)} \|^2} \| h_i \|^2.$$
Where $\Pi_{\mathbf{h}_{\varphi(i)}}\mathbf{h}_i$ represents the orthogonal projection of $\mathbf{h}_i$ onto $\mathbf{h}_{\varphi(i)}$. The norm of $\mathbf{s}_i$ is given by
\[
\|\mathbf{s}_i\| = \|\mathbf{h}_i\| \sqrt{1 + (1 - \theta)(g_i^2 - 2g_i)}.
\] (21)

where $g_i = \rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2 / (1 + \rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2)$. By substituting (21) into (19), we can obtain
\[
\cos^2 \beta_i = \frac{(1 + \rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2 \theta)^2}{(1 + \rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2)^2 (1 + (1 - \theta)(g_i^2 - 2g_i))}.
\] (22)

When $\alpha$ is belongs to the interval $[0, \pi/2)$, the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ is $\alpha + \beta_i$ and the square of the cosine value of $\alpha + \beta_i$ can be expressed as
\[
\cos^2 (\alpha + \beta_i) = (1 - \theta)\cos^2 \beta_i + \theta \sin^2 \beta_i - 2\sqrt{\theta(1 - \theta)} \sin \beta_i \cos \beta_i.
\] (23)

Then, we consider $\beta_i \in [0, \pi/2)$. In this case, by substituting (22) into (23), we can derive
\[
\cos^2 (\alpha + \beta_i) = \frac{1 - \theta}{\rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2 \theta + 2\rho_{\varphi(i)} \| \mathbf{h}_{\varphi(i)} \|^2 \theta + 1}
= \frac{1 - \theta}{\varsigma_i}.
\] (24)

When $\alpha$ is belongs to the interval $[\pi/2, \pi)$, the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ is $\alpha - \beta_i$. Similarly, we have
\[
\cos^2 (\alpha - \beta_i) = \frac{1 - \theta}{\varsigma_i}.
\] (25)

From (24) and (25), we can find that the expressions of $\cos^2 (\alpha + \beta_i)$ with $\alpha \in [0, \pi/2)$ and that of $\cos^2 (\alpha - \beta_i)$ with $\alpha \in [\pi/2, \pi)$ are the same. Therefore, Eq. (16) and (17) can be combined into one equation. We substitute (22) into (16) and (17), and then (24) and (25) into (16) and (17), respectively. Subsequently, by combining (16) and (17), we can obtain (11).

4. Achievable rate region of the adaptive switching method

In this section, we first derive the intersection points of achievable rate region boundaries of NOMA-MRT and MMSE-BF. Based on the intersection points, we proposed the adaptive switching method for the case with two users, which obtains the larger achievable rate region than that derived by employing NOMA-MRT or MMSE-BF only.

4.1 The intersection points of NOMA-MRT’s and MMSE-BF’s rate region boundaries

According to Lemma 1, Lemma 2 and Lemma 3, we can compare the achievable rate regions obtained by NOMA-MRT and MMSE-BF. Then, we can find that neither MMSE-BF nor NOMA-MRT is optimal in all channel states. Consequently, an adaptive switching method is preferred, which can achieve a larger rate region than that derived by MMSE-BF or NOMA-MRT only. Combined with the concept of time-sharing [25], the
The intersection points of NOMA-MRT and MMSE-BF’s rate region boundaries, are given as following:

**Definition 1** If \( r_m(\rho_{n,k}) = \bar{r}_m(\rho_{n,k}) \) and \( r_n(\rho_{n,k}) = \bar{r}_n(\rho_{n,k}), \rho_{n,k} \in [0,\rho], \rho_{n,k} \in [0,\rho] \), we call that the rate region boundaries of NOMA-MRT and MMSE-BF have intersection point at \( \rho_{n,k} \), \( k = 1,2,\cdots,\kappa \), where \( \kappa \) is the number of the intersection points.

The intersection points of NOMA-MRT and MMSE-BF’s rate region boundaries, are given as following:

**Case 1)** When \( \| h_m \| < \| h_n \| \)

By combining (7) with (11), we obtain two equation sets

\[
\begin{align*}
\tilde{\rho}_n \parallel h_n \| ^2 & = \frac{\rho_n \parallel h_n \| ^2(1 + \rho_m \parallel h_m \| ^2 \theta)^2}{\zeta_n + \rho_m \parallel h_m \| ^2 (1 - \theta)} \quad (26) \\
\tilde{\rho}_m \parallel h_m \| ^2 (1 - \theta) & = \frac{\rho_m \parallel h_m \| ^2(1 + \rho_n \parallel h_n \| ^2 \theta)^2}{\zeta_m + \rho_n \parallel h_n \| ^2 (1 - \theta)} \quad (27)
\end{align*}
\]

and

\[
\begin{align*}
\tilde{\rho}_n \parallel h_n \| ^2 & = \frac{\rho_n \parallel h_n \| ^2(1 + \rho_m \parallel h_m \| ^2 \theta)^2}{\zeta_n + \rho_m \parallel h_m \| ^2 (1 - \theta)} \quad (28) \\
\tilde{\rho}_m \parallel h_m \| ^2 (1 - \theta) & = \frac{\rho_m \parallel h_m \| ^2(1 + \rho_n \parallel h_n \| ^2 \theta)^2}{\zeta_m + \rho_n \parallel h_n \| ^2 (1 - \theta)} \quad (29)
\end{align*}
\]

By substituting (26) and (28) into (27) and (29), respectively, the intersection points can be derived by solving the following two equations

\[
c_{1,6} \rho_n^2 + c_{1,5} \rho_n^2 + c_{1,4} \rho_n^3 + c_{1,3} \rho_n^3 + c_{1,2} \rho_n^2 + c_{1,1} \rho_n + c_{1,0} = 0 \quad (30)
\]

where \( c_{1,6} = -\theta^4 \beta_4^2 \), \( c_{1,5} = \theta^3 \beta_4^2 + 3 \theta \beta_3^2, c_{1,4} = -\theta^2 \beta_3^2 \), \( c_{1,3} = -\theta \beta_2^2 \), \( c_{1,2} = \theta \beta_1^2 \), \( c_{1,1} = \theta \beta_0^2 \), \( c_{1,0} = \theta \beta_0^2 \).

According to Eq.(26)(Eq.(28)) with \( \rho_m = \rho - \rho_n \), we can obtain the result

\[
\begin{align*}
\tilde{\rho}_n & = \frac{\rho_n (1 + (\rho - \rho_n) \parallel h_n \| ^2 \theta)^2}{\zeta_n + (\rho - \rho_n) \parallel h_n \| ^2 (1 - \theta)} \quad (32)
\end{align*}
\]
where $c_n = \| h_m \|^4 \theta (\rho - \rho_n) + 2 \| h_m \|^2 \theta (\rho - \rho_n) + 1$. If the rate region boundaries of NOMA-MRT and MMSE-BF have intersection point at $\rho_{n,k} = 1, 2, \cdots, \kappa$, then $\rho_{n,k}$ must satisfy Eq.(32).

For Eq.(30), we consider $\rho_n \in [0, \rho], \bar{\rho}_n \in [0, \rho]$ when $\zeta_{n,m} \leq 0$, or consider $\rho_n \in [0, \rho], \bar{\rho}_n \in (\zeta_{n,m}, \rho]$ when $0 < \zeta_{n,m} < \rho$, as shown in (7). In this case, the intersection point $\rho_{n,k}(k \in \{1, 2, \cdots, \kappa\})$ must satisfy that $\rho_{n,k} \in [0, \rho]$ and $\bar{\rho}_{n,k} \in \{0, \rho\}$ when $\zeta_{n,m} \leq 0$, or $\rho_{n,k} \in [0, \rho]$ and $\bar{\rho}_{n,k} \in (\zeta_{n,m}, \rho]$ when $0 < \zeta_{n,m} < \rho$, where

$$
\bar{\rho}_{n,k} = \frac{\rho_{n,k}(1 + (\rho - \rho_{n,k}) \| h_m \|^2 \theta)^2}{\zeta_{n,k} + (\rho - \rho_{n,k}) \| h_m \|^2 (1 - \theta)}.
$$

Similarly, for Eq.(31), the intersection point $\rho_{n,k}(k \in \{1, 2, \cdots, \kappa\})$ must satisfy that $\rho_{n,k} \in [0, \rho], \bar{\rho}_{n,k} \in [0, \rho]$ when $\zeta_{n,m} \geq \rho$, or $\rho_{n,k} \in [0, \rho], \bar{\rho}_{n,k} \in (0, \zeta_{n,m}]$ when $0 < \zeta_{n,m} < \rho$.

Since deriving the analytic solution of (30) and (31) is not a trivial thing, we use numerical method to derive the solutions.

**Case 2) when $\| h_m \| = \| h_n \| = l$**

By combining (10) with (11), we derive the equation set

$$
\begin{align*}
\bar{\rho}_{n,l} &= \frac{\rho_{n,l}^2 (1 + \rho_{m,l}^2 \theta)^2}{\zeta_n + \rho_{m,l}^2 (1 - \theta)} , \\
\bar{\rho}_{n,l} &= \frac{\rho_{n,l}^2 (1 + \rho_{n,l}^2 \theta)^2}{1 + \rho_{n,l}^2}.
\end{align*}
$$

By substituting (34) into (35), the intersection points can be obtained by solving the following equation

$$
(p_n^2 - \rho_n \rho_n - \frac{2 \rho \theta l^2 + 2 \theta - \rho l^2 + 1}{\theta l^2}) (\rho_n - \rho) (\rho_n + \frac{1}{\theta l^2}) (\rho_n - \rho (\rho_{m,l} + \frac{1}{\theta l^2} + \rho_{m,l}^2 + 1) \rho_n = 0
$$

By solving Eq. (36), we can derive the results

$$
\begin{align*}
\rho_{n,1} &= 0, \quad \rho_{n,2} = \rho , \quad \rho_{n,3} = -\frac{1}{\theta l^2} , \quad \rho_{n,4} = \frac{(\rho l^2 + 1)^2}{\theta l^2} , \\
\rho_{n,(5,6)} &= \frac{\rho (\rho l^2 + 1)^2 + \sqrt{\rho (\rho l^2 + 1)^2 + 8 \rho^2 l^2 - (4 l^2 - 8 \theta + 4)}}{2 \theta l^2 ,}
\end{align*}
$$

where $\rho_{n,k} \in [0, \rho], k = 1, 2, \cdots, 6$.

In the following, we denote by $\bar{\rho}_i (\rho_i)$ the power allocated to user $i$ in the BC with MMSE-BF (NOMA-MRT) scheme, $i = m, n$. Since the transmit power $q_i$ of user $i$ in dual MAC is not equal to the allocated power $\bar{\rho}_i$ for user $i$ in the BC, we should obtain the allocated power $\bar{\rho}_{i,k}$ of user $i$ to design the switching method in the BC, which corresponds to the intersection point $\rho_{i,k} = q_{i,k} / N_0$. According to the duality between MAC and BC,
the SINR at each point on the achievable rate region boundary of BC by performing MMSE-BF is equal to that on the rate region boundary of MAC with MMSE receive filter at the BS, i.e.,

$$\gamma_i^B = \gamma_i^M$$  \hspace{1cm} (38)

We substitute (14) into (4), then substitute (4) and (12) into (38) and replace $p_i$ with $\tilde{p}_i$ by abuse of notation. As a result, $\tilde{p}_i$ can be expressed as

$$\tilde{p}_i = \frac{q_i e_{\phi(i)}(e_i + \rho ||h||^2(1 - \theta))}{e_i e_{\phi(i)} + (1 - \theta)(\rho ||h||^2 e_{\phi(i)} + \rho_{\phi(i)}||h_{\phi(i)}||^2 e_i)}$$  \hspace{1cm} (39)

where $e_i = ||h||^4 \rho_i^2 \theta + 2 ||h||^2 \rho_i \theta + 1$, $e_{\phi(i)} = \rho_{\phi(i)}^2 ||h_{\phi(i)}||^4 \theta + 2 \rho_{\phi(i)} ||h_{\phi(i)}||^2 \theta + 1$, $i = m, n$.

4.2 Adaptive Switching Method to Achieving a Larger Rate Region

Based on the intersection points aforementioned, we develop an adaptive switching method to achieve a larger rate region, i.e., Algorithm 1, described in Table 2, which outputs some parameters such as the user's transmit power and BF vector in any channel condition. Combining with time-sharing, the switching method can achieve a convex hull of the union of NOMA-MRT and MMSE-BF's achievable rate region.

Table 2: Algorithm 1 (Adaptive Switching Method).

| Input: $h_i \in \mathbb{C}^{M \times 1}$ (assuming $||h_m|| < ||h_n||$), $i = m, n$. |
|----------------------------------|
| Output: $p_{m,n}, w_i, i = m,n$. |
| 1. **Step 1:** Obtain the intersection point $\rho_{n,k}$, $k = 1, 2, \cdots, \kappa$. |
| 2. if $||h_m|| < ||h_n||$ |
| 3. Calculate $\rho_{n,k}$ using (30) and (31) |
| 4. else if $||h_m|| = ||h_n||$ |
| 5. Obtain $\rho_{n,k}$ using (37) |
| 6. end if |
| 7. **Step 2:** Define as the mode switching point the intersection point excluding the ends of the interval $[0, \rho]$, and sort the mode switching points in ascending order, i.e., $\rho_{n,1} \leq \cdots \leq \rho_{n,J} \leq \cdots \leq \rho_{n,\Gamma}$, where $\Gamma$ is the number of mode switching points. Divide rate region boundary into $\Gamma + 1$ sections by $\rho_{n,j}$. |
| 8. if $\Gamma = 0$ |
| 9. if (11) is larger compared to (7) or (10), in $[0, \rho]$ |
| 10. Go to Step 3 |
| 11. else |
| 12. Go to Step 4 |
| 13. end if |
| 14. else Divide region boundary (11) into $\Gamma + 1$ sections by $\rho_{n,j}$, $j = 1, 2, \cdots, \Gamma$. |
| 15. for each section between $\rho_{n,j}$ and $\rho_{n,j+1}$, $\rho_{n,0} = 0, \rho_{n,\Gamma+1} = \rho$ |
| 16. if (11) is larger compared to (7) or (10), in this section |
| 17. Go to Step 3 |
| 18. else |
| 19. Go to Step 4 |
| 20. end if |
| 21. end for |
| 22. end if |
| 23. **Step 3:** Applying MMSE-BF |
24. Beamform with $\mathbf{w}_m, \mathbf{w}_n$ described in (14), and transmit $\mathbf{x} = \sqrt{p_m} \mathbf{w}_m s_m + \sqrt{p_n} \mathbf{w}_n s_n$, with $p_{m,n} = [P - \bar{p}_m \bar{p}_n]$, where $\bar{p}_n$ is calculated by (39), with $q_n = \rho_n N_0$, $\rho_m = \rho - \rho_n$ and $\rho_n \in (\rho_{n,j}, \rho_{n,j+1})$.

25. Step 4: Performing NOMA-MRT

26. Beamform with $\mathbf{w}_m, \mathbf{w}_n$ expressed in (6), and transmit $\mathbf{x}$ with $p_{m,n} = [P - p_m p_n]$, where $p_n = \bar{p}_n N_0$, $\bar{p}_n \in (\rho_{n,j}, \rho_{n,j+1})$ and $\rho_{n,j}$ is calculated by (33) (replacing subscript $k$ with subscript $j$).

5. Joint User Pairing and Power Allocation with Transmission Mode

Switching

In this section, for the problem of WSR maximization, we first proposed two optimal power allocation algorithms based on the concept of the achievable rate region, i.e., Power Allocation algorithm for NOMA-MRT mode (NOMA-MRT-PA) and Power Allocation algorithm for MMSE-BF mode (MMSE-BF-PA), by focusing on two-user case. Then, based on two proposed algorithms aforementioned, a Joint User pairing and Power Allocation algorithm (JUPA) is then developed for multi-user case. Finally, a practical transmission method is proposed by combining JUPA with Transmission Mode Switching between NOMA-MRT and MMSE-BF.

5.1 Achieving maximum WSR on the rate region boundary for NOMA-MRT mode

According to [19], the solution to the problem of WSR maximization must be on the boundary of achievable rate region of two-user MISO downlink systems. As a result, when NOMA-MRT mode is employed at BS, the Problem of WSR maximization is formulated as

$$\text{(P2): } \max_{p_{m,n}} U(p_{m,n}) : = \mu_m r_m + \mu_n r_n$$

s.t. $p_{m,n} \in \mathcal{P}_{m,n}$ (40)

where $r_i (i = m,n)$ is the achievable rate of user $i$, defined in Lemma 1 or Lemma 2, and $\mathcal{P}_{m,n} = \{p_{m,n} | 0 \leq p_i \leq p_m + p_\phi(i) = P, i = m,n\}$ is the feasible set of power allocation vector for user $m$ and $n$, in which the power allocation vector corresponding to rate point on the rate region boundary satisfies full power allocation, i.e., $p_i + p_\phi(i) = P$.

For the simplicity of notation, we let $U(\bar{p}_n)$ stand for the WSR $U(p_{m,n})$ in (40) with $p_n = \bar{p}_n N_0$ and $p_m = P - p_n$. The motivations behind proposing NOMA-MRT-PA are as following:

1) According to Lemma 1 and Lemma 2, the expression of the achievable rate region boundary of NOMA-MRT can be given in any channel conditions. Furthermore, the achievable rate $r_i$ is a differentiable or piecewise differentiable function of $\bar{p}_n$ in the interval $[0, \rho]$.

2) The global maximum value of $U(\bar{p}_n)$ in the interval $[0, \rho]$ can be found by selecting the maximum among $U(0), U(\rho)$ and those, corresponding to which the first-order derivative of $U(\bar{p}_n)$ is zeros in the interval $(0, \rho)$. As we know, the local maximum values of $U(\bar{p}_n)$ in the interval $(0, \rho)$ satisfies that the first-order derivative of $U(\bar{p}_n)$ equals to zero.

Theorem 1 When $\| h_m \| < \| h_n \|$, the solution of Problem (P2) is

Case 1) $\xi_{n,m} \leq 0$
where \( p_n^* = \tilde{p}_n^* N_0 p_m^* = P - p_n^* \), \( \tilde{p}_n^* = \arg\max_{\tilde{p}_n} \{ U(0), U(\rho(\tilde{p}_n)) \} \) and

\[
\tilde{p}_{n,1} = \frac{\mu_n \| h_n \|^2 \rho - (1 + \rho \| h_n \|^2)(\mu_n - \mu_m(1 - \theta))}{\mu_n \| h_n \|^2 \theta},
\]

(41)

362 \( \tilde{p}_{n,1} \in [0, \rho] \).

363 **Case 2** \( \zeta_{n,m} \geq \rho \)

364 \( p_{n,m}^* = [p_m^* p_n^*] \).

365 where \( p_n^* = \tilde{p}_n^* N_0 p_m^* = P - p_n^* \), \( \tilde{p}_n^* = \arg\max_{\tilde{p}_n} \{ U(0), U(\rho(\tilde{p}_n)), U(\tilde{p}_{n,2}), U(\tilde{p}_{n,3}) \} \) and

\[
\tilde{p}_{n,(2,3)} = -\frac{\epsilon_1 \pm \sqrt{\epsilon_1^2 - 4\epsilon_2 \epsilon_0}}{2\epsilon_2},
\]

(42)

366 where \( \epsilon_2 = \mu_n \| h_n \|^2 \| h_m \|^4 \| h_n \|^2(1 - \theta) \), \( \epsilon_1 = \| h_n \|^2(\mu_m \epsilon_3 - \mu_n \epsilon_4) \), \( \epsilon_0 = \mu_m \epsilon_3 - \mu_n \| h_n \|^2(1 + \| h_m \|^2 \rho) \), \( \epsilon_3 = \| h_m \|^2(\| h_m \|^2 \rho(1 - \theta) + 1) \), \( \epsilon_4 = \| h_m \|^2(\| h_m \|^2 \rho(1 - \theta) - 2 \theta + 1) \), \( \tilde{p}_{n,2} \in [0, \rho] \) and \( \tilde{p}_{n,3} \in [0, \rho] \).

367 **Case 3** \( 0 < \zeta_{n,m} < \rho \)

368 \( p_{n,m}^* = [p_m^* p_n^*] \).

369 where \( p_n^* = \tilde{p}_n^* N_0 p_m^* = P - p_n^* \), \( \tilde{p}_n^* = \arg\max_{\tilde{p}_n} \{ U(0), U(\rho(\tilde{p}_n)), U(\tilde{p}_{n,1}), U(\tilde{p}_{n,2}), U(\tilde{p}_{n,3}) \} \), \( \tilde{p}_{n,1} = (\zeta_{n,m}, \rho) \), \( \tilde{p}_{n,2} \in [0, \zeta_{n,m}] \) and \( \tilde{p}_{n,3} \in [0, \zeta_{n,m}] \).

371 **Proof** See the Appendix.

372 **Theorem 2** When \( \| h_m \| \leq \| h_n \| \), the solution of Problem (P2) is \( p_{m,n}^* = [P, 0] \) if \( \mu_m \geq \mu_n \). Otherwise, the solution of Problem (P2) is \( p_{m,n}^* = [0, P] \).

375 **Proof** See the Appendix.

376 For the simplicity of description of the proposed NOMA-MRT-PA, we assume \( \| h_m \| \leq \| h_n \| \). By **Theorem 1** and **Theorem 2**, NOMA-MRT-PA is proposed, i.e., Algorithm 2, which is described in Table 3.

377

**Table 3: Algorithm 2 (NOMA-MRT-PA).**

| Input: | \( \mu_i, h_i \in \mathbb{C}^{M \times 1} \) (assuming \( \| h_m \| \leq \| h_n \| \)), \( i = m, n \). |
|-----------------|---------------------------|
| Output: | \( p_{m,n}^* \). |
| 1. | if \( \| h_m \| \leq \| h_n \| \) |
| 2. | Calculate \( \zeta_{n,m} \) using (8) |
| 3. | if \( \zeta_{n,m} \leq 0 \) |
| 4. | \( p_{m,n}^* = [P - \tilde{p}_n^* N_0 \tilde{p}_m^* N_0] \), where \( \tilde{p}_n^* = \arg\max_{\tilde{p}_n} \{ U(0), U(\rho(\tilde{p}_n)), U(\tilde{p}_{n,1}) \} \), \( U(\tilde{p}_{n,1}) \) is calculated by (41), |
5.2 Achieving maximum WSR on rate region boundary for MMSE-BF mode

When MMSE-BF mode is performed at BS, the Problem of WSR maximization can be formulated as

\[(P3): \max_{q_{m,n}} U(q_{m,n}) =: \mu_m \tilde{r}_m + \mu_n \tilde{r}_n \]

s.t. \( q_{m,n} \in Q_{m,n} \) \hspace{1cm} (43)

where \( q_{m,n} = (q_m, q_n) \) is transmit power vector for user \( m \) and \( n \) in the dual MAC, \( \tilde{r}_i \) and \( q_i \) are defined in Lemma 3 and \( Q_{m,n} = \{q_{m,n} | 0 \leq q_i \leq P, q_i + q_{\phi(i)} = P, i = m,n \} \) is the feasible set of transmit power vector for user \( m \) and \( n \).

For the simplicity of notation, we let \( \tilde{r}_m = (\tilde{r}_m) \) stand for the WSR \( (\tilde{r}_m) \) in (43) with \( q_m = \rho_m N_0 \) and \( q_n = P - q_m \). With the similar idea to that, by which Theorem 1 are proposed, we have the following Theorem.

**Theorem 3** The solution of Problem (P3) is \( q^*_{m,n} = [q^*_m, q^*_n] \) in dual two-user MAC, where \( q^*_m = \rho_m N_0 \), \( q^*_n = P - q^*_m \), \( \rho_m = \arg\max_{\rho_m} \{U(0), U(\rho), U(\rho_m, 1), U(\rho_m, 2) \cdots U(\rho_m, 9) \} \) and \( \rho_{m,i} \in [0, \rho] \), which is the roots of the nine-degree equation \( U'(\rho_m) = 0 \), which is given by (45).

**Proof** According to Lemma 3, the WSR \( U(\rho_m) \) can be expressed as

\[ U(\rho_m) = \mu_m \log \left( 1 + \frac{\rho_m \| h_m \|^2 (1 + \rho_n \| h_n \|^2 \theta)^2}{\zeta_m + \rho_m \| h_n \|^2 (1 - \theta)} \right) \]
\[ + \mu_n \log \left( 1 + \frac{\rho_n \| h_n \|^2 (1 + \rho_m \| h_m \|^2 \theta)^2}{\zeta_n + \rho_n \| h_m \|^2 (1 - \theta)} \right) \]

where \( \rho_m = \rho - \rho_m \). In this case, \( U(\rho_m) \) is differentiable function of \( \rho_m \) in the interval \([0, \rho]\). By letting \( U'(\rho_m) = 0 \), we can obtain a nine-degree equation

\[ a_0 \rho_m^9 + a_1 \rho_m^8 + a_2 \rho_m^7 + a_3 \rho_m^6 + a_4 \rho_m^5 + a_5 \rho_m^4 + a_6 \rho_m^3 + a_7 \rho_m^2 + a_8 \rho_m + a_9 = 0 \] \hspace{1cm} (45)

Since the coefficients of Eq. (45) are very complicated, we omit them here. In this case, the global maximum value of \( U(\rho_m) \) in the interval \([0, \rho]\) can be found by selecting the maximum among \( U(0), U(\rho), U(\rho_m, 1), U(\rho_m, 2), \cdots U(\rho_m, 9) \). Consequently, the solution of Problem (P3) is \( q^*_{m,n} = [q^*_m, q^*_n] \) in dual...
two-user MAC, where \( q^*_m = \rho^*_m N_0 \), \( q^*_m = P - q^*_m \), \( \rho^*_m = \arg \max_{\rho_m} \{U(0),U(\rho),U(\rho_{m1}),U(\rho_{m2}),\ldots , U(\rho_{m9})\} \) and \( \rho_{m1} \in [0,\rho] \), \( i = 1,2,\ldots ,9 \). If \( \rho_{m1} \notin [0,\rho] \), \( i = 1,2,\ldots ,9 \), we let \( U(\rho_{m1}) = 0 \).

By Theorem 3, MMSE-BF-PA is proposed, i.e., Algorithm 3, which is described in Table 4.

| Algorithm 3 (MMSE-BF-PA) |
|--------------------------|
| **Input:** \( \mu_i, h_i \in \mathbb{C}^{M \times 1}, i = m,n \). |
| **Output:** \( p^*_{m,n} \). |
| 1. Calculate \( \rho_{m1} \) using (45), \( i = 1,2,\ldots ,9 \). |
| 2. \( \rho^*_m = \arg \max_{\rho_m} \{U(0),U(\rho),U(\rho_{m1}),U(\rho_{m2}),\ldots ,U(\rho_{m9})\} \), where \( \rho_{m1} \in [0,\rho] \). |
| 3. \( p^*_{m,n} = [\bar{p}^*_m,P - \bar{p}^*_m] \), where \( \bar{p}^*_m \) is calculated by (39), with \( q^*_m = \rho^*_m N_0 \). |

5.3 Joint User Pairing and Power Allocation (JUPA) Algorithm

With the aim of maximizing the total sum rate, in the step of forming a user-pair in JUPA, we compare the two sums rates obtained by NOMA-MRT-PA and MMSE-BF-PA, select the transmission mode with the larger sum rate, and then choose the corresponding user-pair to send data. By focusing on the expression of rate region boundary of NOMA-MRT in (7), it is worthwhile noticing the following properties:

1) Considering the original user pairing problem which is how to pair user \( i \) with a user from the set of other unpaired users to obtain the maximum sum rate, we can divide all of unpaired users besides user \( i \) into 3 subsets, i.e., \( \Phi^i_1 \), \( \Phi^i_2 \) and \( \Phi^i_3 \), according to \( \zeta^i \). Particularly, user \( i \) in \( \Phi^i_1 \), \( \Phi^i_2 \) and \( \Phi^i_3 \), satisfies \( \zeta^i \leq 0 \), \( \zeta^i \geq \rho \) and \( 0 < \zeta^i < \rho \), respectively. The original user pairing problem is equivalent to 3 problems, which are to how to pair user \( i \) with a user to obtain the local-maximum sum rate from \( \Phi^i_1 \), \( \Phi^i_2 \) and \( \Phi^i_3 \), respectively. Therefore, the solution of the original user pairing problem is the user-pair, which corresponds to the maximum among the local-maximum sum rates for \( \Phi^i_1 \), \( \Phi^i_2 \) and \( \Phi^i_3 \).

2) Considering the problem which is how to pair user \( i \) with a user from \( \Phi^i_1(\Phi^i_2) \) to achieve local-maximum sum rate, user \( i \) should be paired with the user, whose channel vector can form the minimum (maximum) angle with that of user \( i \), according to (7).

3) For the problem which is how to pair user \( i \) with a user from \( \Phi^i_3 \) to achieve local-maximum sum rate, we divide \( \Phi^i_3 \) into 2 subsets, i.e., \( \Phi^i_{3,1} \) and \( \Phi^i_{3,2} \). The user \( i \) in \( \Phi^i_{3,1} \) and \( \Phi^i_{3,2} \) satisfies \( 0 < \zeta^i \leq \rho/2 \) and \( \rho/2 < \zeta^i < \rho \), respectively. According to (7), the formula \( r^i = \log(1 + \frac{\tilde{\rho}_i}{\|h_i\|^2(1-\theta)}) \) provides dominant contribution to rate region boundary formed by user \( i \) and \( i' \), when \( 0 < \zeta^i \leq \rho/2 \). In this case, for the treatability of the problem, we use the formula \( r^i = \log(1 + \frac{\tilde{\rho}_i}{\|h_i\|^2(1-\theta)}) \) as the expression of \( r^i \) in the whole interval \([0,\rho]\), instead of the piecewise formula as shown in (7). Thus, for the problem which is how to pair user \( i \) with a user from \( \Phi^i_{3,1} \), user \( i \) should be paired with the user whose channel vector can form the minimum angle with that of user \( i \). With the similar analysis, user \( i \) should be paired with the user in \( \Phi^i_{3,2} \) whose channel vector can form the maximum angle with that of user \( i \).

For the simplicity of description of the proposed JUPA (Algorithm 4), described in Table 5, assume that users are ordered as \( \| h_1 \| \leq \cdots \leq \| h_{2K} \| \). Let \( \Phi^i_1(k), \Phi^i_2(k) \) denote the index of \( k \)-th entry in \( \Phi^i_1 \) (\( \Phi^i_2 \)). Besides, let \( \Phi^i_{3,1}(k), \Phi^i_{3,2}(k) \) denote the index of \( k \)-th entry in \( \Phi^i_{3,1} \) (\( \Phi^i_{3,2} \)).

In Algorithm 4, Line 13 (14,15 and 16) is the step for selecting a user to be paired with user \( i \) and deriving the power allocation solution for the obtained user-pair by NOMA-MRT-PA, in the set \( \Phi^i_1, \Phi^i_2, \Phi^i_{3,1} \) and \( \Phi^i_{3,2} \).

Line 18 is the step for similar operation by MMSE-BF-PA. The output of Algorithm 4, \( \Omega \), is termed the pairing
and power allocation configuration, in which \( K \) entries are included and 4 elements are contained in each entry. The first two elements of \( i \)-th entry in \( \Omega \) denote the indexes of the two users in \( i \)-th user-pair, and the third and fourth element stand for power allocation vector and transmission mode adopted by the \( i \)-th user-pair, respectively.

When the number of users is odd, the last remained user can be trivially served by OMA. Specifically, if the channel vector of the last remained user is \( \mathbf{h}_i \), then MRT BF is performed at the BS and the BF vector is given by (6). As a result, the maximum signal to noise ratio can be achieved.

### Table 5: Algorithm 4 (Joint User pairing and Power allocation Algorithm (JUPA)).

| Input: \( \mathbf{h}_i \in \mathbb{C}^{M \times 1} \) (assuming \( \mathbf{h}_1 \| \mathbf{h}_2 \| \ldots \| \mathbf{h}_{2K} \| ) \), \( \mu_i = 1, i = 1, 2, \ldots, 2K. \) |
| Output: \( \Omega \) and \( \mathbf{U} \) |
| Initial: \( \Omega = \emptyset \) and \( \mathbf{U} = \mathbf{0}_{2K \times 2K}. \) |
| 1. for \( i = 1: 2K - 1 \) do |
| 2. if user \( i \) has been paired |
| 3. Go to Line 1 |
| 4. end if |
| 5. for \( j = i + 1: 2K \) do |
| 6. if user \( j \) has been paired |
| 7. Go to Line 5 |
| 8. end if |
| 9. if \( \mathbf{h}_i \) and \( \mathbf{h}_j \) is orthogonal |
| 10. Obtain \( \mathbf{p}_i \) by MMSE-BF-PA, \( \Omega = \Omega \cup \{(i, \phi_1, \mu_1, 0)\} \) and go to Line 1 |
| 11. end if |
| 12. end for |
| 13. Find \( k_1 = \arg \max_{\mathbf{h}} \cos^2 \angle (\mathbf{h}, \mathbf{h}_{\phi_1(k)}) \) and obtain the power allocation solution \( \mathbf{p}_{i,\phi_1(k)} \) for the user-pair \( \{i, \phi_1(k)\} \) by NOMA-MRT-PA |
| 14. Find \( k_2 = \arg \min_{\mathbf{h}} \cos^2 \angle (\mathbf{h}, \mathbf{h}_{\phi_2(k)}) \) and obtain the power allocation solution \( \mathbf{p}_{i,\phi_2(k)} \) for the user-pair \( \{i, \phi_2(k)\} \) by NOMA-MRT-PA |
| 15. Find \( k_{3,1} = \arg \max_{\mathbf{h}} \cos^2 \angle (\mathbf{h}, \mathbf{h}_{\phi_1(k)}) \) and obtain the power allocation solution \( \mathbf{p}_{i,\phi_1(k)} \) for the user-pair \( \{i, \phi_1(k)\} \) by NOMA-MRT-PA |
| 16. Find \( k_{3,2} = \arg \min_{\mathbf{h}} \cos^2 \angle (\mathbf{h}, \mathbf{h}_{\phi_2(k)}) \) and obtain the power allocation solution \( \mathbf{p}_{i,\phi_2(k)} \) for the user-pair \( \{i, \phi_2(k)\} \) by NOMA-MRT-PA |
| 17. Obtain the power allocation vector with the maximum WSR, from the solutions obtained by Line 13-16: |
| \( \mathbf{p}^N = \arg \max_{\mathbf{p}} \left\{ U(\mathbf{p}, \phi_1(k_1)) \right\} \) |
| 18. Let \( \{i_e^N\} \) denote the user-pair which corresponds to \( \mathbf{p}^N. \) |
| 19. Find \( \mathbf{p}_{i_e^N} = \arg \min_{\mathbf{h}} \cos \angle (\mathbf{h}, \mathbf{h}_e) \) and obtain the power allocation solution \( \mathbf{p}_{i_e^N} \) for the user-pair \( \{i_e^N\} \) by MMSE-BF-PA |
| 20. if \( U(\mathbf{p}^N) \geq U(\mathbf{p}_{i_e^N}) \) |
| 21. \( \Omega = \Omega \cup \{(i_e^N, \mathbf{p}, 1)\} \) and \( u_{i_e^N} = u^N_{i_e^N} = 1. \) |
| 22. else |
| 23. \( \Omega = \Omega \cup \{(i_e^M, \mathbf{p}, 0)\} \) and \( u_{i_e^M} = u^M_{i_e^M} = 1. \) |
| 24. end if |
| 25. end for |
1) Complexity Analysis:

The computational complexity of the proposed JUPA mainly comes from the computation of the angle between channels of two users and from solving the nine-degree Eq. (45), by which the optimal power allocation scheme is obtained for MMSE-BF mode.

Note that the computational complexity for computing the angle between channels of two users is \( O(M) \). In addition, we use upper-case letter ‘ \( C \) ’ to indicate the computational complexity for solving the nine-degree Eq.(45). AS the number of computing the angle is \( 2K - 2(i - 1) - 1 \) in \( i \)-th outer ‘for’ loop in Algorithm 4, for MMSE-BF mode or NOMA-MRT mode, the computational complexity for the user pairing process is \( O((2K/2)^2M) \). On the other hand, the computational complexity for the power allocation process is \( O((2K/2)^2C) \). Consequently, the overall computational complexity of Algorithm 4 is \( O((2K-1)!(2K/2)^2M) + O((2K-1)!(2K/2)^2C) \).

For exhaustive research based scheme, i.e., Exhaustive Search based user Pairing and Power Allocation (ES-PPA) described in Section 6.2, the number of all possible pairing schemes is \( (2K-1)! \). In each pairing scheme, there are \( K \) user-pairs. Then, for each user-pair, the angle between channels of two users and the allocated power of two users for MMSE-BF mode should be calculated. Therefore, the total computational complexity of Exhaustive Research based scheme, is \( O((2K-1)!((2K/2)^2M) + O((2K-1)!((2K/2)^2C)) \).

5.4 JUPA/TMS: A Practical Transmission Method

By combining JUPA with Transmission Mode Switching (TMS) between NOMA-MRT and MMSE-BF, a practical transmission scheme termed as JUPA/TMS is proposed. JUPA/TMS, i.e., Algorithm 5, is described in Table 6, where \( \beta \) is control parameter for transmission mode switching and \( \Omega(i) \) denotes the \( j \)-th element in the \( i \)-th entry in the set \( \Omega \). Noting that TDMA(FDMA) is adopted in Algorithm 5 (Line10), it is efficient when \( M \) is small. When the scenario with sufficient number of antennas at base station is considered, e.g., \( M \geq 2K \), the performance can be further improved by combining the proposed JUPA with SDMA, which is beyond the scope of this paper.

Table 6: Algorithm 5 (JUPA/TMS).

| Input: \( h_i \in \mathbb{C}^{M \times 1}, s_i, i = 1,2,\ldots,2K \). |
| Output: \( x_1, x_2,\ldots, x_K \). |
| 1. Perform JUPA to obtain pairing and power allocation configuration \( \Omega \). |
| 2. for \( i = 1:K \) do |
| 3. \( (m_n, p_{m_n}, T_{m_n}) \leftarrow (\Omega(i)_1, \Omega(i)_2, \Omega(i)_3, \Omega(i)_4) \). |
| 4. if \( T_{m_n} = 1 \) |
| 5. Calculate \( \mathbf{w}_m \) and \( \mathbf{w}_n \) with (6). |
| 6. else if \( T_{m_n} = 0 \) |
| 7. Calculate \( \mathbf{w}_m \) and \( \mathbf{w}_n \) with (14). |
| 8. end if |
| 9. \( x_i \leftarrow \sqrt{p_m} \mathbf{w}_ms_m + \sqrt{p_n} \mathbf{w}_ns_n \). |
| 10. Transmit \( x_i \) in \( i \)-th frequency/time slot |
| 11. end for |

6. Results and Discussion

In this section, we first provide numerical results on the rate regions of different schemes (ZFBB, MRT, MMSE-BF and NOMA-MRT) under various channel conditions and validate the performance of the proposed Adaptive Switching Method by comparing with other schemes. Then, we demonstrate the optimality of NOMA-MRT-PA.
and MMSE-BF-PA. Furthermore, we also verify the performance of the proposed JPUA by comparing with traditional transmission schemes in simulation.

6.1 The Adaptive Switching Method

In Figure 3, numerical results of performance comparison among NOMA-MRT, MMSE-BF, MRT and ZFBF, are derived in various channel state. When \( \| \mathbf{h}_n \| = 2 \) and \( \alpha = \pi/8 \) implying that the angle between two users' channel vectors is smaller in space, NOMA-MRT gets better performance than MMSE-BF absolutely and the rate region of MMSE-BF is completely within that of NOMA-MRT as illustrated in (a). In (b), when \( \| \mathbf{h}_n \| = 2 \) and \( \alpha = \pi/4 \), larger \( R_n \) can be obtained by implementing NOMA-MRT in high \( R_m \) area, while in low \( R_m \) area larger \( R_n \) can be derived by using MMSE-BF. When \( \| \mathbf{h}_n \| = 2 \) and \( \alpha = 3\pi/8 \) which implies that the angle between channel vectors of two users becomes larger, MMSE-BF performs better than NOMA-MRT completely as shown in (c). In symmetric case where \( \| \mathbf{h}_n \| = \| \mathbf{h}_n \| = 1 \) and \( \alpha = \pi/4 \), the rate region of MMSE-BF covers that of NOMA-MRT. From (a) to (d), as MMSE-BF can optimally trade off fighting interference to other users and the background Gaussian noise, MMSE-BF get better performance than ZFBF and MRT absolutely, in any channel state.

![Figure 3: Rate region comparison of ZFBF, MRT, MMSE-BF and NOMA-MRT in various channel state.](image)

As demonstrated in Figure 4, the adaptive switching method can obtain a larger rate region than the other schemes, which can be illustrated as the convex hull the union of NOMA-MRT and MMSE-BF's rate regions by combining with time-sharing.
Figure 4: Rate region of the Adaptive Switching Method with time-sharing. $P/N_0 = 10$dB, $\|h_m\| = 1$, $\|h_n\| = 2$, $\alpha = \pi/4$. Also plotted for comparison are rate regions of ZFBF, MRT, MMSE-BF and NOMA-MRT, without time-sharing.

6.2 Joint User Pairing and Power Allocation with Transmission Mode Switching

In Figure 5, with different weight vectors, the rate points on rate region boundary and the corresponding maximum-WSR points are obtained by NOMA-MRT-PA, MMSE-BF-PA and exhaustive search, respectively. When weight vector $\mathbf{u}$ is $(0.5,0.5)$, $(0.33,0.67)$ or $(0.25,0.75)$, rate points obtained by NOMA-MRT-PA or MMSE-BF-PA, are superposed together with that derived by exhaustive research, which means that NOMA-MRT-PA and MMSE-BF can derive optimal rate point for different weight vector as shown in (a). The rate points obtained by NOMA-MRT-PA for the case $\mathbf{u} = (0.33,0.67)$ and $\mathbf{u} = (0.25,0.75)$ are superposed together, since the rate points to achieve the maximum-WSR for $\mathbf{u} = (0.33,0.67)$ and $\mathbf{u} = (0.25,0.75)$ both occurs at the inflection point of the achievable rate region boundary of NOMA-MRT. In (b), with $\mathbf{u} = (0.5,0.5)$, the WRSs obtained by NOMA-MRT (red curve) and MMSE-BF (blue curve) are given respectively, when the allocated power of user $m$ varies, with $P$ being normalized to 1. The maximum-WSR points corresponding to the rate point with $\mathbf{u} = (0.5,0.5)$ in (a) can be obtained by NOMA-MRT-PA and MMSE-BF-PA, respectively, each of which is also superposed with that derived by exhaustive research. (c) and (d) are the cases for $\mathbf{u} = (0.33,0.67)$ and $\mathbf{u} = (0.25,0.75)$, respectively.
Figure 5: Rate points and the corresponding maximum-WSR points obtained by NOMA-MRT-PA, MMSE-BF-PA and exhaustive research. $P/N_0 = 10\text{ dB}$, $\parallel \mathbf{h}_m \parallel = 2$, $\parallel \mathbf{h}_n \parallel = 1$, $\alpha = \pi/4$. (a) rate points under different weight vectors. (b), (c) and (d) is for maximum-WSR point with $\mathbf{u} = (0.5,0.5), \mathbf{u} = (0.33,0.67)$ and $\mathbf{u} = (0.25,0.75)$, respectively.

To validate the effectiveness of the proposed JUPA scheme, other schemes which combine the existed user pairing algorithms (Exhaustive Search, Greedy Algorithm, CORrelation-based-pairing[27], [28] and RANdom pairing Algorithm) with NOMA-MRT-PA and MMSE-BF-PA, are also considered for comparison in simulation, including Exhaustive Search based user Pairing and Power Allocation algorithm (ES-PPA), Greedy Algorithm based user Pairing and Power Allocation algorithm (GA-PPA), CORrelation-based-pairing based user Pairing and Power Allocation algorithm (COR-PPA) and RANdom pairing based user Pairing and Power Allocation algorithm (RAN-PPA). In the following, we assume noise power $N_0 = -10$ dB, the cell radius $R = 30$ m, the path loss exponent $\alpha = 3$ and TDMA technology is adopted. The bandwidth is normalized to one. Assume that the weights of all users are set to one. The key idea of GA-PPA is to pair the two users with the maximum sum rate obtained by NOMA-MRT-PA or MMSE-BF-PA. Mathematically, for a fixed user $i$, we pair it with user $j^*$, if

$$j^* = \arg\max_j \max_i U_{ij}^{\text{NOMA-MRT}}, U_{ij}^{\text{MMSE-BF}}, j \neq i$$

where $U_{ij}^{\text{NOMA-MRT}}$ and $U_{ij}^{\text{MMSE-BF}}$ stand for the sum rate of user $i$ and $j$, obtained by NOMA-MRT-PA and MMSE-BF-PA, respectively. With the similar architecture to the outer 'for' loop in the Algorithm 4, the numbers of computing the angle between channels of two users used by NOMA-MRT-PA and the number of solving the Eq.(45) used by MMSE-BF-PA are both $2K - 2(i - 1) - 1$ in $i$-th outer 'for' loop for GA-PPA. Therefore, its computational complexity is $O((2K/2)^2 M) + O((2K/2)^3 C)$ in total. In RAN-PPA and COR-PPA, we first obtain user pairing solution by RANdom pairing algorithm and CORrelation-based-pairing...
algorithm, respectively, and then derive the sum rate for each user-pair by selecting the larger one among the
sum rates achieved by NOMA-MRT-PA and MMSE-BF-PA.

The key idea of CORrelation-based-pairing algorithm is only to pair two users with the maximum channel
correlation. Specifically, for a fixed user $i$, we pair it with user $j^*$ satisfying

$$j^* = \arg\max_j \frac{|h_i^H h_j|^2}{\|h_i\|^2 \|h_j\|^2}, j \neq i.$$  

Since the ideas of forming a user-pair in COR-PPA and the proposed JUPA are both based on the angle
between channels of the two users, COR-PPA and JUPA have the same computational complexity (considering
the complexity of computing the angle between two users’ channels and solving the Eq.(45), i.e.,
$O((2K/2)^2 M) + O((2K/2) C)$. In ES-PPA, we first obtain $(2K - 1)!!$ possible pairing schemes by exhaustive
research and then select the pairing scheme with maximum total sum rate, which is the sum of sum rates of $K$
user-pairs.

For summary, JUPA and COR-PPA have roughly the same computational complexity. However, GA-PPA and
ES-PPA require much higher computational complexity than JUPA, especially for ES-PPA. Due to the random
pairing method adopted in RAN-PPA, its computational complexity is the lowest, which only results from
computing the angle between two users’ channels used by NOMA-MRT-PA and from solving the Eq.(45)
utilized by MMSE-BF-PA, for each user-pair. The computational complexity of these schemes above are
summarized as in Table 7.

| User Pairing and Power Allocation schemes | Computation complexity |
|-----------------------------------------|-----------------------|
| GA-PPA                                  | $O((2K/2)^2 M) + O((2K/2)^2 C)$ |
| COR-PPA                                 | $O((2K/2)^2 M) + O((2K/2) C)$ |
| RAN-PPA                                 | $O((2K/2) M) + O((2K/2) C)$ |
| ES-PPA                                  | $O((2K - 1)!!(2K/2) M) + O((2K - 1)!!(2K/2) C)$ |
| JUPA                                    | $O((2K/2)^2 M) + O((2K/2) C)$ |

In Figure 6 and Figure 7, the sum rates of different schemes, including ES-PPA, RAN-PPA, GA-PPA, COR-
PPA and the proposed JUPA, are compared, when the transmit power $P$ and the number of antennas at the BS
varies, respectively. In Figure 6, the sum rates of all schemes grow with the transmit power $P$ increasing.

Compared with ES-PPA using exhaustive search to obtain the user pairing solution, the proposed JUPA only
suffers very slight performance loss. However, the computational complexity of JUPA is only
$O((2K/2)^2 M) + O((2K/2) C)$, while ES-PPA is $O((2K - 1)!!(2K/2) M) + O((2K - 1)!!(2K/2) C)$.

Further, JUPA outperforms GA-PPA, since GA does not consider the performance loss caused
by the users with poor channels. COR-PPA performs worse than RAN-PPA, since correlation-based user
pairing algorithm breaks the channels orthogonality, i.e., the users with orthogonal channel are never paired
together.
As demonstrated in Figure 7, it is can be observed that the proposed JUPA results in a slight performance loss compared with ES-PPA. However, the JUPA outperforms other schemes and the performance gain over other schemes decreases as the number of antennas at the BS increases.
Conclusion

In this paper, the design of adaptive TMS between NOMA-MRT and MMSE-BF to maximize the sum rate for
downlink MISO systems, was investigated. Firstly, the closed-form expressions of the boundary of achievable
rate region for NOMA-MRT and MMSE-BF, were obtained. It has been shown that when the channel vectors of
the two users are greatly correlated, the achievable rate region of NOMA-MRT include that of MMSE-BF
completely. However, when the two channels are almost orthogonal, the opposite conclusion can be drawn. As a
result, an adaptive switching method is developed to achieve a larger rate region for the two-user case.
Subsequently, the optimal power allocation algorithms to maximize weighted sum rate for both NOMA-MRT
mode and MMSE-BF mode were presented. Finally, the low-complexity JUPA is consequently developed to
further improve sum-rate performance for the multi-user case. Compared with the exhaustive research based
scheme with the computational complexity of $O((2K - 1)!((2K/2)M) + O((2K - 1)!((2K/2)C))$, the
proposed JUPA can obtain a much lower complexity of $O((2K/2)^2M) + O((2K/2)C)$ and only suffers a
slight sum-rate performance loss, whereas outperforms other conventional schemes.

Appendix

Proof of Lemma 1

Case 1) $\theta > 0$

When the equality holds for (5), the achievable rate region boundary of NOMA-base BF for MISO broadcast
channel can be achieved. As a result, the achievable rate region boundary of NOMA-MRT can be given by

\[
\begin{align*}
\{ r_n &= \log(1 + \bar{\rho}_n \| h_n \|^2), \\
\{ r_m &= \min\{ \log(1 + \gamma^B_{m,n}), \log(1 + \gamma^B_m) \},
\end{align*}
\]

where

\[
\begin{align*}
\gamma^B_{m,n} &= \frac{\bar{\rho}_m \| h_n \|^2(1 - \theta)}{1 + \bar{\rho}_m \| h_n \|^2}, \\
\gamma^B_m &= \frac{\bar{\rho}_m \| h_m \|^2}{1 + \bar{\rho}_m \| h_m \|^2(1 - \theta)},
\end{align*}
\]

and $\bar{\rho}_m = \rho - \bar{\rho}_n$.

Define

\[
f(\bar{\rho}_n) = \gamma^B_{m,n} - \gamma^B_m.
\]

We let $f(\bar{\rho}_n) = 0$. Subsequently, we can derive a quadratic equation with respect to $\bar{\rho}_n$

\[
m_1\bar{\rho}_n^2 + m_2\bar{\rho}_n + m_3 = 0
\]

where $m_1 = \| h_n \|^2 \| h_m \|^2(2\theta - \theta^2)$, $m_2 = -\left( \| h_n \|^2(1 - \theta) - \| h_m \|^2 \right) - \rho \| h_n \|^2 \| h_m \|^2(2\theta - \theta^2)$, $m_3 = \rho \| h_n \|^2(1 - \theta) - \| h_m \|^2$.

As $\theta > 0$, we solve Eq. (50) with quadratic formula. The discriminant of the Eq. (50) can be expressed as

26
\[ \Delta = m_1^2 - 4m_1m_3 = (\rho \parallel \mathbf{h}_m \parallel^2 \parallel \mathbf{h}_m \parallel^2 (2\theta - \theta^2) - (\parallel \mathbf{h}_n \parallel^2 (1 - \theta) - \parallel \mathbf{h}_m \parallel^2))^2. \]

As \( \Delta \geq 0 \), the Eq. (50) has two roots both of which are real numbers and the roots are given by

\[ \tilde{\rho}_{n,1}^{(\Delta)} = \rho, \]
\[ \tilde{\rho}_{n,2}^{(\Delta)} = \zeta_{n,m}, \] 

where \( \zeta_{n,m} = \frac{\mathbf{h}_n^T (1 - \theta) - \mathbf{h}_m^T}{\mathbf{h}_n^T \mathbf{h}_m^T (2\theta - \theta^2)} \). When \( \Delta = 0 \), \( \tilde{\rho}_{n,1}^{(\Delta)} = \tilde{\rho}_{n,2}^{(\Delta)}. \)

Assuming \( m_1 > 0 \), we have the following cases.

Case a) When \( \rho \leq \zeta_{n,m} \), i.e. \( \tilde{\rho}_{n,1}^{(\Delta)} \leq \tilde{\rho}_{n,2}^{(\Delta)} \), the function \( f(\tilde{\rho}_n) \geq 0 \) with \( \tilde{\rho}_n \in [0, \rho] \), which means that \( \log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_{m,n}^B) \) with \( \tilde{\rho}_n \in [0, \rho] \). In this case, \( r_m = \log(1 + \gamma_{m,n}^B) \) according to (47). When \( \tilde{\rho}_n = \rho \), the equality in \( f(\tilde{\rho}_n) \geq 0 \) holds.

Case b) When \( \zeta_{n,m} \leq 0 \), i.e., \( \tilde{\rho}_{n,2}^{(\Delta)} \leq \tilde{\rho}_{n,1}^{(\Delta)} \), the function \( f(\tilde{\rho}_n) \leq 0 \) with \( \tilde{\rho}_n \in [0, \rho] \), which means that \( \log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_{m,n}^B) \) with \( \tilde{\rho}_n \in [0, \rho] \). In this case, \( r_m = \log(1 + \gamma_{m,n}^B) \) according to (47). When \( \zeta_{n,m} = 0 \), and \( \tilde{\rho}_n = 0 \) or \( \tilde{\rho}_n = \rho \), the equality in \( f(\tilde{\rho}_n) \geq 0 \) holds.

Case c) When \( 0 < \tilde{\rho}_{n,2}^{(\Delta)} = \zeta_{n,m} < \rho \), the function \( f(\tilde{\rho}_n) \geq 0 \) with \( \tilde{\rho}_n \in [0, \zeta_{n,m}] \) and \( f(\tilde{\rho}_n) \leq 0 \) with \( \tilde{\rho}_n \in (\zeta_{n,m}, \rho] \), which means that \( \log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_{m,n}^B) \) with \( \tilde{\rho}_n \in [0, \zeta_{n,m}] \) and \( \log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_{m,n}^B) \) with \( \tilde{\rho}_n \in (\zeta_{n,m}, \rho] \).

By combining Case a), Case b) and Case c), we can derive (7).

Case 2) \( \theta = 0 \)

As \( \theta = 0 \), we can obtain the result

\[ \left\{ \begin{array}{l}
\gamma_{m,n}^B = \frac{\tilde{\rho}_m \parallel \mathbf{h}_m \parallel^2}{1 + \tilde{\rho}_n \parallel \mathbf{h}_n \parallel^2}, \\
\gamma_{m}^T = \frac{\tilde{\rho}_m \parallel \mathbf{h}_m \parallel^2}{1 + \tilde{\rho}_n \parallel \mathbf{h}_n \parallel^2}.
\end{array} \right. \]

In this case, due to \( \parallel \mathbf{h}_m \parallel \ll \parallel \mathbf{h}_n \parallel, f(\tilde{\rho}_n) > 0 \). As a result, we can derive (9).

Proof of Lemma 2

The upper bound \( r_i^{\text{upper}} = \log(1 + P i^2/N_0) \) of achievable rate of user \( i \), can be derived by allocating all the power \( P \) to user \( i \) with the other user's rate being zero. Then, we obtain two extreme rate tuple \( (r_m^{\text{upper}}, 0) \), \( (0, r_{m,2}^{\text{upper}}) \). As a result, when \( \parallel \mathbf{h}_m \parallel = \parallel \mathbf{h}_n \parallel = l \), by time-sharing between the two extreme rate tuple, the achievable rate region boundary of MISO broadcast channel can be expressed as [22]
\[ r_{TS}^n = \lambda \log \left( 1 + \frac{P l^2}{N_0} \right), \]
\[ r_{TS}^m = (1 - \lambda) \log \left( 1 + \frac{P l^2}{N_0} \right), \]  
(52)

where \( \lambda \in [0,1] \) represents the fraction of the time allocated to user \( n \). We consider the case in which SIC is executed at user \( n \). According to (46) and (47), when \( \|h_m\| = \|h_n\| = l \), the achievable rate region boundary can be characterized as

\[
\begin{cases}
    r_n = \log(1 + \frac{P l^2}{N_0}), \\
    r_m = \log(1 + \frac{P l^2 \cos^2 \alpha}{N_0 + P_m l^2}).
\end{cases}
\]  
(53)

\[ \forall t \in (0,\infty), \alpha \in [0,\pi), \lambda \in [0,1], p_m \in [0,P] \text{ and } p_n \in [0,P], \text{ we let} \]
\[ r_{TS}^n = r_n. \]  
(54)

We define
\[ h(p_m) = r_{TS}^m - r_m. \]  
(55)

By substituting (54) into (55), we can get the result
\[ h(p_m) = \log \left( \frac{N_0 + P l^2}{N_0 + P l^2 - l^2 P_m (1 - \cos^2 \alpha)} \right) \geq 0 \]  
(56)

From (56), when \( \cos^2 \alpha = 1 \), (53) is equivalent to (52). Consequently, we obtain (10). When SIC is executed at user \( m \), the same result can be achieved.

**Proof of Theorem 1**

**Case 1) \( \zeta_{n,m} \leq 0 \)**

According to Lemma 1, when \( \zeta_{m,n} \leq 0 \), the WSR \( U(\tilde{p}_n) \) can be expressed as
\[ U(\tilde{p}_n) = \mu_n \log(1 + \tilde{p}_n \|h_n\|^2) + \mu_n \log \left( \frac{1 + \tilde{p}_m \|h_m\|^2(1 - \theta)}{1 + \tilde{p}_n \|h_n\|^2} \right), \]  
(57)

where \( \tilde{p}_m = \rho - \tilde{p}_n \). In this case, \( U(\tilde{p}_n) \) is differentiable function of \( \tilde{p}_n \) in the interval \([0,\rho] \). By letting \( U'(\tilde{p}_n) = 0 \), we can obtain a linear equation
\[ \varepsilon_1 \tilde{p}_n + \varepsilon_0 = 0 \]
(58)

where \( \varepsilon_1 = \mu_n \|h_n\|^2 \theta, \varepsilon_0 = (1 + \rho \|h_n\|^2)(\mu_n - \mu_m(1 - \theta)) - \mu_n \|h_n\|^2 \rho \theta \). As a result, we can obtain (41). In this case, the global maximum value of \( U(\tilde{p}_n) \) in the interval \([0,\rho] \) can be found by selecting the
maximum among \(U(0), U(\rho)\) and \(U(\tilde{\rho}_{n,1})\). Therefore, the solution of Problem (P2) is \(p_{m,n}^* = [p_m^* p_n^*]\), where \(p_n^* = \tilde{\rho}_n N_0, p_m^* = P - p_n^*, \tilde{\rho}_n = \arg\max_{\rho} \{U(0), U(\rho), U(\tilde{\rho}_{n,1})\} \).

**Case 2**) \(\zeta_{n,m} \geq \rho\)

According to Lemma 1, when \(\zeta_{n,m} \geq \rho\), the WSR \(U(\tilde{\rho}_n)\) can be expressed as

\[
U(\tilde{\rho}_n) = \mu_n \log(1 + \tilde{\rho}_n \| h_n \|^2) + \mu_m \log(1 + \frac{\tilde{\rho}_m \| h_m \|^2}{1 + \tilde{\rho}_n \| h_m \|^2(1 - \theta)}).
\]

where \(\tilde{\rho}_m = \rho - \tilde{\rho}_n\). By letting \(U'(\tilde{\rho}_n) = 0\), we can obtain a quadratic equation

\[
\epsilon_2 \tilde{\rho}_m^2 + \epsilon_1 \tilde{\rho}_m + \epsilon_0 = 0
\]

**Case 3**) \(0 < \zeta_{n,m} < \rho\)

When \(0 \leq \tilde{\rho}_n \leq \zeta_{n,m}\), \(U(\tilde{\rho}_n)\) can be expressed as \(58\) according to Lemma 1. Therefore, the power allocation vector for the global maximum value of \(U(\tilde{\rho}_n)\) in the interval \([0, \zeta_{n,m}]\) is the one corresponding to the maximum among \(U(0), U(\zeta_{n,m}), U(\tilde{\rho}_{n,2})\) and \(U(\tilde{\rho}_{n,3})\), where \(\tilde{\rho}_{n,2} \in [0, \zeta_{n,m}], \tilde{\rho}_{n,3} \in [0, \zeta_{n,m}]\).

When \(\zeta_{n,m} < \tilde{\rho}_n \leq \rho\), \(U(\tilde{\rho}_n)\) can be expressed as \(57\). Therefore, the power allocation vector for the global maximum value of \(U(\tilde{\rho}_n)\) in the interval \((\zeta_{n,m}, \rho]\) is the one corresponding to the maximum among \(U(\zeta_{n,m}), U(\rho)\) and \(U(\tilde{\rho}_{n,1})\), where \(\tilde{\rho}_{n,1} \in (\zeta_{n,m}, \rho]\).

We combine the two cases above, in which \(0 \leq \tilde{\rho}_n \leq \zeta_{n,m}\) and \(\zeta_{n,m} < \tilde{\rho}_n \leq \rho\). Consequently, in the whole interval \([0, \rho]\), the solution of Problem (P2) is \(p_{m,n}^* = [p_m^* p_n^*]\), where \(p_n^* = \tilde{\rho}_n N_0, p_m^* = P - p_n^*, \tilde{\rho}_n = \arg\max_{\rho} \{U(0), U(\rho), U(\tilde{\rho}_{n,1}), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}, \tilde{\rho}_{n,1} \in (\zeta_{n,m}, \rho], \tilde{\rho}_{n,2} \in [0, \zeta_{n,m}]\) and \(\tilde{\rho}_{n,3} \in [0, \zeta_{n,m}]\).

**Proof of Theorem 2**

If \(\mu_m \geq \mu_n\), the WSR \(\mu_m r_m + \mu_n r_n \leq \mu_m (r_m + r_n)\). According to Lemma 2, \(r_m + r_n = \log(1 + \rho^2)\). Consequently, if \(\mu_m \geq \mu_n\), \(\mu_m r_m + \mu_n r_n \leq \mu_m \log(1 + \rho^2)\), i.e., \(U(0) = \mu_m \log(1 + \rho^2)\) is the global maximum value of \(U(\tilde{\rho}_n)\) in the interval \([0, \rho]\), implying that the solution of Problem (P2) is \(p_{m,n}^* = [P, 0]\). If \(\mu_m < \mu_n\), the proof is similar to that of the above case, in which \(\mu_m \geq \mu_n\).

**Data Availability**

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

**Conflicts of Interest**

The authors declare that there is no conflict of interest regarding the publication of this paper.
Funding Statement

This work is supported by the Non-funded science and technology project of Zhanjiang [grant number 2020B01067,2020B01037,2019B01076]; the Special Foundation of Talents Project at Lingnan Normal University [grant number ZL2048]; the Fundamental Research Funds project for provincial universities in Heilongjiang Province under [grant number Hkdxp201804].

Supplementary Materials

No Supplementary Materials are provided in this manuscript.

References

1. Le, N.-T., Tran, L.-N., Vu, Q.-D., Jayalath, D.: Energy-efficient resource allocation for OFDMA heterogeneous networks. IEEE Transactions on Communications 67(10), 7043–7057 (2019)
2. Zhao, Z., Jia, M., Guo, Q.: Resource allocation algorithm based on bandwidth demands for OFDMA systems with imperfect channel quality information. Sensors & Transducers 167(3), 151 (2014).
3. Zeng, J., Lv, T., Liu, R.P., Su, X., Peng, M., Wang, C., Mei, J.: Investigation on evolving single-carrier NOMA into multi-carrier NOMA in 5G. IEEE Access 6, 48268–48288 (2018)
4. Islam, S.R., Avazov, N., Dobre, O.A., Kwak, K.-S.: Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. IEEE Communications Surveys & Tutorials 19(2), 721–742 (2016)
5. Liu, G., Wang, Z., Hu, J., Ding, Z., Fan, P.: Cooperative NOMA broadcasting/multicasting for low-latency and high-reliability 5G cellular V2X communications. IEEE Internet of Things Journal 6(5), 7828–7838 (2019)
6. Makki, B., Chitti, K., Behravan, A., Alouini, M.-S.: A survey of NOMA: Current status and open research challenges. IEEE Open Journal of the Communications Society 1, 179–189 (2020)
7. Cover, T.M., Thomas, J.A.: Elements of Information Theory [Internet]. John Wiley & Sons, New York (2012)
8. Hanif, M.F., Ding, Z., Ratnarajah, T., Karagiannidis, G.K.: A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems. IEEE Transactions on Signal Processing 64(1), 76–88 (2015)
9. Chen, C., Cai, W., Cheng, X., Yang, L., Jin, Y.: Low complexity beamforming and user selection schemes for 5G MIMO-NOMA systems. IEEE Journal on Selected Areas in Communications 35(12), 2708–2722 (2017)
10. Amin, S.H., Mehana, A.H., Soliman, S.S., Fahmy, Y.A.: User capacity in downlink MISO-NOMA systems. In: 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–7 (2018). IEEE
[11] Sun, Q., Han, S., Xu, Z., Wang, S., Chih-Lin, I., Pan, Z.: Sum rate optimization for MIMO non-orthogonal multiple access systems. In: 2015 IEEE Wireless Communications and Networking Conference (WCNC), pp. 747–752 (2015). IEEE

[12] Zhang, Q., Li, Q., Qin, J.: Robust beamforming for nonorthogonal multiple-access systems in MISO channels. IEEE Transactions on Vehicular Technology 65(12), 10231–10236 (2016)

[13] Chen, Z., Ding, Z., Xu, P., Dai, X.: Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink. IEEE Communications Letters 20(6), 1263–1266 (2016)

[14] Chen, Z., Ding, Z., Dai, X., Karagiannidis, G.K.: On the application of quasi-degradation to MISO-NOMA downlink. IEEE Transactions on Signal Processing 64(23), 6174–6189 (2016)

[15] Ding Z., Fan P., Poor HV. Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. IEEE Transactions on Vehicular Technology 65(8):6010–6023 (2015)

[16] Nguyen HV, Nguyen VD, Dobre OA, Nguyen DN, Dutkiewicz E, Shin OS. Joint power control and user association for NOMA-based full-duplex systems. IEEE Transactions on Communications 67(11):8037–8055 (2019)

[17] Nguyen TV, Nguyen VD, da Costa DB, An B. Hybrid user pairing for spectral and energy efficiencies in multiuser MISO-NOMA networks with SWIPT. IEEE Transactions on Communications 68(8):4874–4890 (2020)

[18] Jorswieck, E.A., Larsson, E.G.: Linear precoding in multiple antenna broadcast channels: Efficient computation of the achievable rate region. In: 2008 International ITG Workshop on Smart Antennas, pp. 21–28 (2008). IEEE

[19] Zhao, Z., Chen, W.: An adaptive switching method for sum rate maximization in downlink MISO-NOMA systems. In: GLOBECOM 2017-2017 IEEE Global Communications Conference, pp. 1–6 (2017). IEEE

[20] Jorswieck, E.A., Larsson, E.G., Danev, D.: Complete characterization of the pareto boundary for the MISO interference channel. IEEE Transactions on Signal Processing 56(10), 5292-5296 (2008)

[21] Jorswieck, E.A., Larsson, E.G.: Monotonic optimization framework for the two-user MISO interference channel. IEEE Transactions on Communications 58(7), 2159–2168 (2010)

[22] Shang, X., Chen, B., Poor, H.V.: Multiuser MISO interference channels with single-user detection: Optimality of beamforming and the achievable rate region. IEEE Transactions on Information Theory 57(7), 4255–4273 (2011)

[23] Ding, Z., Liu, Y., Choi, J., Sun, Q., Elkashlan, M., Chih-Lin, I., Poor, H.V.: Application of non-orthogonal multiple access in LTE and 5G networks. IEEE Communications Magazine 55(2), 185–191 (2017)

[24] Sklar, B.: Rayleigh fading channels in mobile digital communication systems. I. characterization. IEEE Communications magazine 35(7), 90–100 (1997)

[25] Tse, D., Viswanath, P.: Fundamentals of Wireless Communication. Cambridge university press, Cambridge (2005)
[26] Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge university press, Cambridge (2012)

[27] Tseng SM, Chen YF, Fang HH. Cross PHY/APP layer user grouping and power allocation for uplink multiantenna NOMA video communication systems. IEEE Systems Journal 14(3):3351-3359 (2019)

[28] Nguyen HV, Kim HM, Kang GM, Nguyen KH, Bui VP, Shin OS. A Survey on Non-Orthogonal Multiple Access: From the Perspective of Spectral Efficiency and Energy Efficiency. Energies 13(16):4106 (2020)
Figures

Figure 1

System model with transmission mode switching between NOMA-MRT and MMSE-BF

Figure 2

The geometric description of BF vectors for ZFBF, MRT and MMSE-BF.
Figure 3

Rate region comparison of ZFBF, MRT, MMSE-BF and NOMA-MRT in various channel state.
Figure 4
Rate region of the Adaptive Switching Method with time-sharing

Figure 5
Rate points and the corresponding maximum-WSR points obtained by NOMA-MRT-PA, MMSE-BF-PA and exhaustive research
Figure 6

Sum rate versus $P$ for different transmission schemes.
Figure 7

Sum rate versus the number of the antennas for several transmission schemes.