RESEARCH

Adaptive transmission mode switching in downlink MISO-NOMA systems: sum rate maximization and the achievable rate region

Zhixin Zhao^{1*}, Dong Wang², Hongwei Zhang¹ and Haitao Sang¹

*Correspondence:

Zhixin.Z@lingnan.edu.cn

¹School of Information

Engineering, Lingnan Normal

University, Zhanjiang, China

Full list of author information is available at the end of the article

Abstract

In this paper, the design of adaptive transmission mode switching (TMS) to maximize the sum rate for the downlink multiple-input-single-output based non-orthogonal multiple access (MISO-NOMA) systems is investigated. Firstly, the closed-form expressions of the boundary of achievable rate region of two candidate transmission mode, i.e., NOMA-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 mode, 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 achievable rate region, the optimal solutions to the problem of WSR optimization for NOMA-MRT and MMSE-BF are obtained, respectively. Subsequently, by exploiting the optimal solutions aforementioned for two transmission modes and the high efficiency of TMS, a low-complexity Joint 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 computation 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.

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

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 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 sufficient 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, owing 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 computation 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], in [14], it was 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 (ZF-BF).

Due to the equivalence between the maximization of the weighted sum rate and the acquisition of the maximum weighted sum rate (WSR) point on achievable rate region for the MISO downlink, the problems for characterization of the achievable rate region for MISO broadcast channel [15, 16] and for MISO interference channel [17–19], were investigated, respectively. Specifically, in [15], 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 [16], 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 [17]. 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 [18]. In [19], 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-NOMA systems. In this paper, from the perspective of the achievable rate region, we focus on the design of adaptive transmission mode switching (TMS) 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 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 equivalence between the maximization of the weighted sum rate and the acquisition of the maximum weighted sum rate point on achievable rate region for the MISO downlink, 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 transmission mode switching between NOMA-MRT and MMSE-BF, a practical transmission method is developed. By making use of the closed-expression of the optimal solution by NOMA-MRT-PA(MMSE-BF-PA), JUPA can be performed with a low computation complexity $\mathcal{O}(K^2M)$, where $2K$ is the number of users and M stands for the number of antennas at the BS, while the computation complexity of the Exhaustive Search based user Pairing and Power Allocation algorithm (ES-PPA) is $\mathcal{O}((2K-1)!!KM)$, 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 NOMA-based MISO downlink transmission schemes. 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. The numerical results is 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 $\|\cdot\|$. $\mathbf{x} \in \mathbb{C}^{M \times 1}$ means that $\mathbf{x}$ is an $M \times 1$ complex vector. $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex Gaussian random variable with mean μ and variance σ^2 . $\langle \mathbf{x}, \mathbf{y} \rangle$ 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. $2K$ 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 perform SIC to cancel the co-channel interference. However, with the number of users in a group growing, the processing complexity and delay at users dramatically increase [20]. As a result, we consider only grouping two users and adopting user pairing scheme to decrease processing complexity and delay in the NOMA downlink. Assume that user m and n are paired over the shared spectrum. The observation at user i is given by

$$y_i = \mathbf{h}_i^H \mathbf{x} + z_i, \quad i = m, n, \quad (1)$$

where $\mathbf{h}_i \sim \mathcal{CN}(0, \sigma_i^2 \mathbf{I}_M)$ is the channel vector from BS to user i , $z_i \sim \mathcal{CN}(0, N_0)$ is the additive white Gaussian noise (AWGN) at user i and N_0 is the corresponding noise power. For the channel vector $\mathbf{h}_i$, the variance of the channel from BS to user i is modeled as [14], [21]

$$\sigma_i^2 = \begin{cases} d_i^{-\alpha'}, & \text{if } d_i > d_0 \\ d_0^{-\alpha'}, & \text{otherwise,} \end{cases}$$

where d_i denotes the distance between BS and user i , α' denotes the path loss exponent and the parameter d_0 avoids the singularity when d_i is small. Furthermore, $\mathbf{x} = \sqrt{p_m} \mathbf{w}_m s_n + \sqrt{p_n} \mathbf{w}_n s_n$ is the signal transmitted by the BS, where s_i and $\mathbf{w}_i$

are the scalar signal and normalized BF vector for user i , respectively. p_i is the transmit power allocated to user i . Assume that the BS can obtain perfect CSI.

In this paper, we attempt to maximize the sum rate of users by jointly optimizing the pairing relationship matrix $\mathbf{U}$ as well as power allocation vector $\mathbf{p}$ and BF matrix $\mathbf{W}$ of each user-pair. Mathematically, the optimization problem can be formulated as

$$(P1) : \max_{\mathbf{p}_{m,n}, \mathbf{W}_{m,n}, \mathbf{U}} \sum_{m=1}^{2K} \sum_{n=m+1}^{2K} u_{m,n} (R_m + R_n) \quad (2a)$$

$$\text{s.t. } u_{m,n} \in \{0, 1\}, 1 \leq m, n \leq 2K, \quad (2b)$$

$$u_{m,n} = u_{n,m}, 1 \leq m, n \leq 2K, \quad (2c)$$

$$u_{m,m} = 0, 1 \leq m \leq 2K, \quad (2d)$$

$$\sum_{m=1}^{2K} u_{m,n} = 1, 1 \leq n \leq 2K \quad (2e)$$

$$\sum_{n=1}^{2K} u_{m,n} = 1, 1 \leq m \leq 2K \quad (2f)$$

$$0 \leq u_{m,n} (p_m + p_n) \leq P, 1 \leq m, n \leq 2K \quad (2g)$$

$$\|\mathbf{w}_m\| = \|\mathbf{w}_n\| = 1, 1 \leq m, n \leq 2K, \quad (2h)$$

where R_i is the rate of user i , $i = m, n$. $\mathbf{U}$ is the user pairing relationship matrix with $2K$ -dimension, in which the m -th row and n -th column element is denoted by $u_{m,n}$. $\mathbf{p}_{m,n} = [p_m, p_n]$ and $\mathbf{W}_{m,n} = [\mathbf{w}_m, \mathbf{w}_n]$ are power allocation vector and BF matrix of the user-pair $\{m, n\}$. As shown in (2b), $u_{m,n} = 1$ represents that user m and n are paired together and constitute a user-pair. Otherwise, user m and n are not paired together. (2c) and (2d) imply that $\mathbf{U}$ is a symmetric matrix, and the diagonal elements are zero as a user can't be paired with itself. (2e) and (2f) indicate that a user can be paired with only one user. Furthermore, (2g) represents the total power allocated to a user-pair $\{m, n\}$, which is upper bounded to P . (2h) 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 rate R_i of user i with the equality holding in (3) or (5). Therefore, for the treatability of the problem (P1), we select the BF scheme 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 $\mathbf{U}$, and power allocation vector $\mathbf{p}$ and BF matrix $\mathbf{W}$ 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 $\mathbf{w}_i$, the achievable rate region of the MISO broadcast channel (BC) is given by the set of rate tuples $(\tilde{R}_m, \tilde{R}_n)$ satisfying

$$R_i \leq \log(1 + \gamma_i^B), \quad (3)$$

where

$$\gamma_i^B = \frac{p_i |\mathbf{h}_i^H \mathbf{w}_i|^2}{N_0 + p_{\varphi(i)} |\mathbf{h}_i^H \mathbf{w}_{\varphi(i)}|^2}, \quad (4)$$

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 $\varphi(i)$ 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 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. Specifically, for a fixed pair of users m and n , the user n with strong channel conditions, first decodes s_m and subtracts this from its received signal y_n . As a result, user n can decode s_n without the interference caused by user m . However, user m simply treat s_n as noise and decodes its own signal s_m . Therefore, for a fixed normalized BF vector $\mathbf{w}_i$, when SIC is carried out at user n , 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

$$\begin{cases} R_n \leq \log(1 + \frac{p_n |\mathbf{h}_n^H \mathbf{w}_n|^2}{N_0}) \\ R_m \leq \min\{\log(1 + \gamma_{m,n}^B), \log(1 + \gamma_m^B)\}, \end{cases} \quad (5)$$

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.,

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

In this paper, the angle between $\mathbf{h}_m$ and $\mathbf{h}_n$ is denoted by α as, where

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

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

2.4 The equivalence between maximizing weighted sum rate and enlarging achievable rate region

For a given weight vector, the solution of optimization problem of maximizing the weighted sum rate (WSR) must be on the boundary of achievable rate region of the two-user MISO broadcast channel [16]. Thus, maximizing the weighted sum rate

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

$$\mathbf{w}_i^{\text{MRT}} = \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|}, \quad i = m, n. \quad (6)$$

NOMA-MRT is the combination of NOMA and MRT.

Lemma 1 When $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as

Case 1) $\theta > 0$

$$r_n = \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2),$$

$$r_m = \begin{cases} \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2}\right), & \zeta_{n,m} \leq 0 \text{ or } \zeta_{n,m} < \tilde{\rho}_n \leq \rho \\ \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)}\right), & \text{otherwise,} \end{cases} \quad (7)$$

where

$$\zeta_{n,m} = \frac{\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2}{\|\mathbf{h}_m\|^2 \|\mathbf{h}_n\|^2 (2\theta - \theta^2)}, \quad (8)$$

$\theta = \sin^2 \alpha$, $\rho = P/N_0$ and $\tilde{\rho}_i = p_i/N_0$, $i = m, n$.

Case 2) $\theta = 0$

$$\begin{cases} r_n = \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2), \\ r_m = \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2}\right), \end{cases} \quad (9)$$

Proof See the Appendix.

Lemma 2 When $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, the achievable rate region boundary of NOMA-MRT for MISO broadcast channel can be expressed as

$$\begin{cases} r_n = \log\left(1 + \frac{p_n l^2}{N_0}\right) \\ r_m = \log\left(1 + \frac{p_m l^2}{N_0 + p_n l^2}\right). \end{cases} \quad (10)$$

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) [22]. 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 Fig. 1.

By MAC-BC duality theory [22], 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 Theorem.

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 $(\tilde{r}_m, \tilde{r}_n)$ on the achievable rate region boundary of BC satisfies

$$\tilde{r}_i = \log(1 + \gamma_i^M), \quad i = m, n, \quad (11)$$

where

$$\gamma_i^M = \frac{\rho_i \|\mathbf{h}_i\|^2 (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta)^2}{\varsigma_i + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 (1 - \theta)}, \quad (12)$$

$\varsigma_i = \rho_{\varphi(i)}^2 \|\mathbf{h}_{\varphi(i)}\|^4 \theta + 2\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta + 1$, $\rho_i = q_i/N_0$ 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 [22], the normalized BF vectors of user i with MMSE-BF in MISO broadcast channel is given by

$$\mathbf{w}_i^{\text{MMSE}} = \frac{(N_0 \mathbf{I} + q_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H)^{-1} \mathbf{h}_i}{\|(N_0 \mathbf{I} + q_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H)^{-1} \mathbf{h}_i\|}. \quad (13)$$

By the matrix inversion Lemma [23], Eq.(13) can be written as

$$\mathbf{w}_i^{\text{MMSE}} = \frac{\left(\mathbf{I} - \frac{\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \right) \mathbf{h}_i}{\left\| \left(\mathbf{I} - \frac{\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \right) \mathbf{h}_i \right\|}. \quad (14)$$

The SINR γ_i^M of user i in dual MAC with MMSE receiver filter have the following form

$$\gamma_i^M = \frac{q_i |\mathbf{h}_i^H \mathbf{w}_i^{\text{MMSE}}|^2}{N_0 + q_{\varphi(i)} |\mathbf{h}_{\varphi(i)}^H \mathbf{w}_i^{\text{MMSE}}|^2}. \quad (15)$$

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

$$\gamma_i^M = \begin{cases} \frac{q_i \|\mathbf{h}_i\|^2 \cos^2 \beta_i}{N_0 + q_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \cos^2(\alpha + \beta_i)}, & \alpha \in [0, \pi/2) \\ \frac{q_i \|\mathbf{h}_i\|^2 \cos^2 \beta_i}{N_0 + q_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \cos^2(\alpha - \beta_i)}, & \alpha \in [\pi/2, \pi). \end{cases} \quad (16)$$

where β_i represent the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_i$, $\alpha + \beta_i$ is the angle between $\mathbf{w}_i^{\text{MMSE}}$ and $\mathbf{h}_{\varphi(i)}$ when $\alpha \in [0, \pi/2)$ and $\alpha - \beta_i$ is that when $\alpha \in [\pi/2, \pi)$, as illustrated in Fig. 1.

The square of the cosine value of β_i can be expressed as

$$\cos^2 \beta_i = \frac{|\langle \mathbf{w}_i^{\text{MMSE}}, \mathbf{h}_i \rangle|^2}{\|\mathbf{h}_i\|^2} \quad (18)$$

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

$$\cos^2 \beta_i = \frac{\|\mathbf{h}_i\|^2 (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta)^2}{\|\mathbf{s}_i\|^2 (1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2)^2}, \quad (19)$$

where

$$\mathbf{s}_i = \left(\mathbf{I} - \frac{\rho_{\varphi(i)} \mathbf{h}_{\varphi(i)} \mathbf{h}_{\varphi(i)}^H}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \right) \mathbf{h}_i.$$

In according to the definition of inner product, $\mathbf{s}_i$ can be written as

$$\begin{aligned} \mathbf{s}_i &= \mathbf{h}_i - \frac{\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \frac{\langle \mathbf{h}_i, \mathbf{h}_{\varphi(i)} \rangle \mathbf{h}_{\varphi(i)}}{\|\mathbf{h}_{\varphi(i)}\|^2} \\ &= \mathbf{h}_i - \frac{\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2}{1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2} \Pi_{\mathbf{h}_{\varphi(i)}} \mathbf{h}_i \end{aligned} \quad (20)$$

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

$$\|\mathbf{s}_i\| = \|\mathbf{h}_i\| \sqrt{1 + (1 - \theta)(g_i^2 - 2g_i)}. \quad (21)$$

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

$$\cos^2 \beta_i = \frac{(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta)^2}{(1 + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2)^2 (1 + (1 - \theta)(g_i^2 - 2g_i))}. \quad (22)$$

When α 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

$$\cos^2(\alpha + \beta_i) = (1 - \theta) \cos^2 \beta_i + \theta \sin^2 \beta_i - 2\sqrt{\theta(1 - \theta)} \sin \beta_i \cos \beta_i \quad (23)$$

Then, we consider $\beta_i \in [0, \pi/2)$. In this case, by substituting (22) into (23), we can derive

$$\begin{aligned} \cos^2(\alpha + \beta_i) &= \frac{1 - \theta}{\rho_{\varphi(i)}^2 \|\mathbf{h}_{\varphi(i)}\|^4 \theta + 2\rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 \theta + 1} \\ &= \frac{1 - \theta}{\varsigma_i} \end{aligned} \quad (24)$$

When α 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

$$\cos^2(\alpha - \beta_i) = \frac{1 - \theta}{\varsigma_i} \quad (25)$$

From (24) and (25), we can find that the expressions of $\cos^2(\alpha + \beta_i)$ with $\alpha \in [0, \pi/2)$ and $\cos^2(\alpha - \beta_i)$ with $\alpha \in [\pi/2, \pi)$ are the same in both cases. Therefore, Eqs.(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). $\square$

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 [22], the adaptive switching method can achieve a convex hull of the union of MMSE-BF and NOMA-MRT's achievable rate region.

Definition 1 If $r_m(\tilde{\rho}_{n,k}) = \tilde{r}_m(\rho_{n,k})$ and $r_n(\tilde{\rho}_{n,k}) = \tilde{r}_n(\rho_{n,k})$, $\tilde{\rho}_{n,k} \in [0, \rho]$, $\rho_{n,k} \in [0, \rho]$, we call that the rate region boundaries of NOMA-MRT and MMSE-BF have intersection point at $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$, where κ 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 $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$

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

$$\begin{cases} \tilde{\rho}_n \|\mathbf{h}_n\|^2 = \frac{\rho_n \|\mathbf{h}_n\|^2 (1 + \rho_m \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)}, \\ \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2} = \frac{\rho_m \|\mathbf{h}_m\|^2 (1 + \rho_n \|\mathbf{h}_n\|^2 \theta)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)} \end{cases} \quad (26)$$

$$\begin{cases} \tilde{\rho}_n \|\mathbf{h}_n\|^2 = \frac{\rho_n \|\mathbf{h}_n\|^2 (1 + \rho_m \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)}, \\ \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2} = \frac{\rho_m \|\mathbf{h}_m\|^2 (1 + \rho_n \|\mathbf{h}_n\|^2 \theta)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)} \end{cases} \quad (27)$$

and

$$\begin{cases} \tilde{\rho}_n \|\mathbf{h}_n\|^2 = \frac{\rho_n \|\mathbf{h}_n\|^2 (1 + \rho_m \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)} \\ \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)} = \frac{\rho_m \|\mathbf{h}_m\|^2 (1 + \rho_n \|\mathbf{h}_n\|^2 \theta)^2}{\varsigma_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)}. \end{cases} \quad (28)$$

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

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

where $c_{1,6} = -\theta^4 l_n^3 l_m^3$, $c_{1,5} = \theta^3 l_n^2 l_m^2 (3l_n - 3l_m - \theta l_n + 3\rho\theta l_n l_m)$, $c_{1,4} = -\theta^2 l_n l_m (3\rho^2 \theta^2 l_n^2 l_m^2 - 2\rho\theta^2 l_n^2 l_m + 5\rho\theta l_n^2 l_m - 9\rho\theta l_n l_m^2 + \rho l_n^2 l_m + \theta^2 l_n l_m - 2\theta l_n^2 - \theta l_n l_m + 3l_n^2 - 6l_n l_m + 3l_m^2)$, $c_{1,3} = -\theta(-\rho^3 \theta^3 l_n^3 l_m^3 + \rho^2 \theta^3 l_n^3 l_m^2 - \rho^2 \theta^2 l_n^3 l_m^2 + 9\rho^2 \theta^2 l_n^2 l_m^3 - 2\rho^2 \theta^3 l_n^2 l_m^2 - 2\rho\theta^3 l_n^2 l_m^2 + 3\rho\theta^2 l_n^3 l_m + \rho\theta^2 l_n^2 l_m^2 - 3\rho\theta l_n^3 l_m + 12\rho\theta l_n^2 l_m^2 - 9\rho\theta l_n l_m^3 - \rho l_n^3 l_m + \rho l_n^2 l_m^2 - 2\theta^2 l_n^2 l_m + \theta^2 l_n l_m^2 + \theta l_n^3 + \theta l_n^2 l_m - 3\theta l_n l_m^2 - l_n^3 + 4l_n^2 l_m - 4l_n l_m^2 + l_m^3)$, $c_{1,2} = -\rho^2 \theta^3 l_n^2 l_m^2 - \rho^2 \theta^2 l_n^3 l_m v_m - 3\rho^2 \theta^2 l_n^2 l_m^2 v_m + 3\rho^2 \theta^2 l_n l_m^3 + \rho^2 \theta l_n^3 l_m v_m - \rho^2 \theta l_n^2 l_m^2 + \rho\theta^3 l_n l_m - \rho\theta^2 l_n^3 v_m + \rho\theta^2 l_n l_m^2 + \rho\theta l_n^3 v_m - 2\rho\theta l_n^2 l_m v_m - \rho\theta l_n^2 l_m + 6\rho\theta l_n l_m^2 v_m + 2\rho\theta l_n l_m^2 - 2\rho\theta l_n^3 - \rho l_n^2 l_m v_m + \theta^2 l_n^2 v_m^2 - 2\theta^2 l_n l_m v_m + 4\theta l_n l_m v_m - \theta l_m^2 - l_n^2 v_m^2 + l_n l_m v_m - l_m^2 v_m$, $c_{1,1} = 2\rho\theta l_n l_m v_m v_n - \rho\theta^2 l_n l_m v_n + \rho\theta l_n l_m v_n - \rho\theta l_n l_m v_m + \rho\theta^2 l_n^2 v_m v_n + \rho^2 \theta l_n l_m^2 v_m + 2\rho l_n l_m v_m - \rho l_n^2 v_m v_n - \rho\theta l_m^2 v_n - \theta l_n v_m^2 - \rho l_m^2 v_m - l_m v_m v_n l_n v_m^2$, $c_{1,0} = \rho v_m (\rho l_m + 1) (l_m - l_n + \theta l_n)$, $v_m = \rho\theta l_m + 1$, $v_n = \rho l_m + 1$.

$$c_{2,6}\rho_n^6 + c_{2,5}\rho_n^5 + c_{2,4}\rho_n^4 + c_{2,3}\rho_n^3 + c_{2,2}\rho_n^2 + c_{2,1}\rho_n + c_{2,0} = 0 \quad (31)$$

where $c_{2,6} = -\theta^4 (1 - \theta) l_n^2 l_m^3$, $c_{2,5} = \theta^3 l_n l_m^2 (1 - \theta) (2l_n - 2l_m + 3\rho\theta l_n l_m)$, $c_{2,4} = -\theta^2 (1 - \theta) l_m (3\rho^2 \theta^2 l_n^2 l_m^2 + 4\rho\theta l_n^2 l_m - 6\rho\theta l_n l_m^2 + \rho l_n^2 l_m + 2l_n^2 - 3l_n l_m + l_m^2)$, $c_{2,3} = \theta (1 - \theta) (\rho^3 \theta^3 l_n^3 l_m^3 + 2\rho^2 \theta^2 l_n^2 l_m^2 - 6\rho\theta^2 l_n l_m^3 + 2\rho^2 \theta l_n^2 l_m^2 + 3\rho\theta l_n^2 l_m - 6\rho\theta l_n l_m^2 + 3\rho\theta l_m^3 + \rho l_n^2 l_m - \rho l_n l_m^2 + l_n^2 - 2l_n l_m + l_m^2)$, $c_{2,2} = (1 - \theta) (2\rho^3 \theta^3 l_n^3 l_m^3 - \rho^3 \theta^2 l_n^2 l_m^2 - \rho^2 \theta^2 l_n^2 l_m + 3\rho^2 \theta^2 l_n l_m^2 - 3\rho^2 \theta^2 l_m^3 - \rho^2 \theta l_n^2 l_m + 2\rho^2 \theta l_n l_m^2 - \rho\theta l_n^2 + 3\rho\theta l_n l_m - 2\rho\theta l_m^2 + \rho l_n l_m + l_n)$, $c_{2,1} = \rho (1 - \theta) (\rho\theta l_m + 1) (\rho\theta l_m^2 - \rho l_n l_m - l_n)$, $c_{2,0} = 0$.

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

$$\tilde{\rho}_n = \frac{\rho_n (1 + (\rho - \rho_n) \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_n + (\rho - \rho_n) \|\mathbf{h}_m\|^2 (1 - \theta)} \quad (32)$$

where $\varsigma_n = \|\mathbf{h}_m\|^4 \theta (\rho - \rho_n)^2 + 2\|\mathbf{h}_m\|^2 \theta (\rho - \rho_n) + 1$. If the rate region boundaries of NOMA-MRT and MMSE-BF have intersection point at $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$, then $\rho_{n,k}$ must satisfy Eq.(32).

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

$$\tilde{\rho}_{n,k} = \frac{\rho_{n,k} (1 + (\rho - \rho_{n,k}) \|\mathbf{h}_m\|^2 \theta)^2}{\varsigma_{n,k} + (\rho - \rho_{n,k}) \|\mathbf{h}_m\|^2 (1 - \theta)} \quad (33)$$

$$\varsigma_{n,k} = \|\mathbf{h}_m\|^4 \theta (\rho - \rho_{n,k})^2 + 2\|\mathbf{h}_m\|^2 \theta (\rho - \rho_{n,k}) + 1.$$

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

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 $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$

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

$$\begin{cases} \tilde{\rho}_n l^2 = \frac{\rho_m l^2 (1 + \rho_n l^2 \theta)^2}{\varsigma_m + \rho_n l^2 (1 - \theta)} \end{cases} \quad (34)$$

$$\begin{cases} \frac{\tilde{\rho}_n l^2}{1 + \tilde{\rho}_m l^2} = \frac{\rho_n l^2 (1 + \rho_m l^2 \theta)^2}{\varsigma_n + \rho_m l^2 (1 - \theta)}. \end{cases} \quad (35)$$

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

$$(\rho_n^2 - \rho \rho_n - \frac{2\rho\theta l^2 + 2\theta - \rho l^2 + 1}{\theta^2 l^4})(\rho_n - \rho)(\rho_n + \frac{1}{\theta l^2})(\rho_n - \frac{\rho\theta l^2 + 1}{\theta l^2})\rho_n = 0 \quad (36)$$

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

$$\begin{aligned} \rho_{n,1} &= 0, & \rho_{n,2} &= \rho \\ \rho_{n,3} &= -\frac{1}{\theta l^2}, & \rho_{n,4} &= \frac{(\rho\theta l^2 + 1)}{\theta l^2} \\ \rho_{n,(5,6)} &= \frac{\rho\theta l^2 \pm \sqrt{\rho^2 \theta^2 l^4 + 8\rho\theta l^2 - (4\rho l^2 - 8\theta + 4)}}{2\theta l^2} \end{aligned} \quad (37)$$

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

We denote by $\tilde{p}_i$ ($i = m, n$) the power allocated to user i in the BC with MMSE-BF scheme. Since the transmit power q_i of user i in dual MAC is not equal to the allocated power $\tilde{p}_i$ for user i in the BC, we should obtain the allocated power $\tilde{p}_{i,k}$ of user i to design the switching method in the BC, which corresponds to the mode switching 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.,

$$\gamma_i^B = \gamma_i^M \quad (38)$$

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

$$\tilde{p}_i = \frac{q_i e_{\varphi(i)} (e_i + \rho \|\mathbf{h}_i\|^2 (1 - \theta))}{e_i e_{\varphi(i)} + (1 - \theta) (\rho_i \|\mathbf{h}_i\|^2 e_{\varphi(i)} + \rho_{\varphi(i)} \|\mathbf{h}_{\varphi(i)}\|^2 e_i)} \quad (39)$$

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

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, described in Algorithm 1, 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.

5 Joint User Pairing and Power Allocation with Transmission Mode Switching

In this section, for the problem of weighted sum rate (WSR) maximization, we first proposed two optimal power allocation algorithms based on the concept of 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, 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 rate region boundary for NOMA-MRT mode

According to [16], the solution to the problem of maximizing the WSR 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 WSR maximization Problem is formulated as

$$\begin{aligned} \text{(P2)} : \max_{\mathbf{p}_{m,n}} U(\mathbf{p}_{m,n}) &:= \mu_m r_m + \mu_n r_n \\ \text{subject to } \mathbf{p}_{m,n} &\in \mathcal{P}_{m,n} \end{aligned} \quad (40)$$

where r_i ($i = m, n$) is the achievable rate of user i , defined in Lemma 1 or Lemma 2, and $\mathcal{P}_{m,n} = \{\mathbf{p}_{m,n} | 0 \leq p_i \leq P, p_i + p_{\varphi(i)} = P, i = m, n\}$ 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., $p_i + p_{\varphi(i)} = P$.

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

1) According to Lemma 1 and 2, the expression of 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 $\tilde{\rho}_m$ in the interval $[0, \rho]$.

2) The global maximum value of $U(\tilde{\rho}_n)$ 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 $U(\tilde{\rho}_n)$ is zeros in the interval $(0, \rho)$. As we know, the local maximum values of $U(\tilde{\rho}_n)$ in the interval $(0, \rho)$ satisfies that the first-order derivative of $U(\tilde{\rho}_n)$ equals to zero.

Theorem 1 When $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$, the solution of Problem (P2) is

Algorithm 1 Adaptive Switching Method

Input: $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ (assuming $\|\mathbf{h}_m\| \leq \|\mathbf{h}_n\|$), $i = m, n$.

Output: $\mathbf{p}_{m,n}, \mathbf{w}_i, i = m, n$.

Step 1: Obtain the intersection point $\rho_{n,k}$, $k = 1, 2, \dots, \kappa$.

If $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$

Calculate $\rho_{n,k}$ using (30) and (31)

Else $\|\mathbf{h}_m\| = \|\mathbf{h}_n\|$

Calculate $\rho_{n,k}$ using (37)

Step 2: Define as the mode switching point the intersection point excluding the end of the interval $[0, \rho]$, and sort the mode switching points in ascending order, i.e., $\rho_{n,1} \leq \dots \leq \rho_{n,j} \leq \dots \leq \rho_{n,\Gamma}$, where Γ is the number of mode switching points. Divide rate region boundary into $\Gamma + 1$ sections by $\rho_{n,j}$.

If $\Gamma = 0$

If (11) is larger in $[0, \rho]$, go to **Step 3**

Else go to **Step 4**

Else Divide region boundary (11) into $\Gamma + 1$ sections by $\rho_{n,j}$, $j = 1, 2, \dots, \Gamma$.

For each section between $\rho_{n,j}$ and $\rho_{n,j+1}$ ($\rho_{n,0} = 0, \rho_{n,\Gamma+1} = \rho$)

If (11) is larger in this section, go to **Step 3**

Else go to **Step 4**

Step 3 Applying MMSE-BF

Beamform with $\mathbf{w}_m, \mathbf{w}_n$ described in (14), and transmit $\mathbf{x} = \sqrt{p_m} \mathbf{w}_m s_n + \sqrt{p_n} \mathbf{w}_n s_n$, with $\mathbf{p}_{m,n} = [P - \tilde{p}_n, \tilde{p}_n]$, where $\tilde{p}_n$ is calculated by (39), with $q_n = \rho_n N_0$, $\rho_m = \rho - \rho_n$ and $\rho_n \in (\rho_{n,j}, \rho_{n,j+1})$.

Step 4 Performing NOMA-MRT

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

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

the power allocation vector, which corresponds to the maximum among $U(0)$, $U(\rho)$ and $U(\tilde{\rho}_{n,1})$, where

$$\tilde{\rho}_{n,1} = \frac{\mu_n \|\mathbf{h}_n\|^2 \rho \theta - (1 + \rho \|\mathbf{h}_n\|^2)(\mu_n - \mu_m(1 - \theta))}{\mu_n \|\mathbf{h}_n\|^2 \theta}, \quad (41)$$

and $\tilde{\rho}_{n,1} \in [0, \rho]$.

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

the power allocation vector, which corresponds to the maximum among $U(0)$, $U(\rho)$, $U(\tilde{\rho}_{n,2})$ and $U(\tilde{\rho}_{n,3})$, where

$$\tilde{\rho}_{n,(2,3)} = \frac{-\epsilon_1 \pm \sqrt{\epsilon_1^2 - 4\epsilon_2\epsilon_0}}{2\epsilon_2}, \quad (42)$$

where $\epsilon_2 = \mu_n \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^4 \theta(1 - \theta)$, $\epsilon_1 = \|\mathbf{h}_n\|^2(\mu_m \epsilon_3 - \mu_n \epsilon_4)$, $\epsilon_0 = \mu_m \epsilon_3 - \mu_n \|\mathbf{h}_n\|^2(1 + \|\mathbf{h}_m\|^2 \rho)$, $\epsilon_3 = \|\mathbf{h}_m\|^2(\|\mathbf{h}_m\|^2 \rho(1 - \theta) + 1)$ and $\epsilon_4 = \|\mathbf{h}_m\|^2(\|\mathbf{h}_m\|^2 \rho(1 - \theta) - 2\theta + 1)$, $\tilde{\rho}_{n,2} \in [0, \rho]$ and $\tilde{\rho}_{n,3} \in [0, \rho]$.

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

the power allocation vector, which corresponds to the maximum among $U(0)$, $U(\rho)$, $U(\zeta_{n,m})$, $U(\tilde{\rho}_{n,1})$ and $U(\tilde{\rho}_{n,2})$ and $U(\tilde{\rho}_{n,3})$, where $\tilde{\rho}_{n,1} \in (\zeta_{n,m}, \rho]$, $\tilde{\rho}_{n,2} \in [0, \zeta_{n,m}]$ and $\tilde{\rho}_{n,3} \in [0, \zeta_{n,m}]$.

Proof See the Appendix.

Theorem 2 When $\|\mathbf{h}_m\| = \|\mathbf{h}_n\|$, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [P, 0]$ if $\mu_m \geq \mu_n$. Otherwise, the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [0, P]$.

Proof See the Appendix.

For the simplicity of description of the proposed NOMA-MRT-PA, we assume $\|\mathbf{h}_m\| \leq \|\mathbf{h}_n\|$. By Theorem 1 and Theorem 2, NOMA-MRT-PA is proposed, which is described in Algorithm 2.

5.2 Achieving maximum WSR on rate region boundary for MMSE-BF mode

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

$$\begin{aligned} \text{(P3)} : \max_{\mathbf{q}_{m,n}} U(\mathbf{q}_{m,n}) &:= \mu_m \tilde{r}_m + \mu_n \tilde{r}_n \\ \text{subject to } \mathbf{q}_{m,n} &\in \mathcal{Q}_{m,n} \end{aligned} \quad (43)$$

where $\mathbf{q}_{m,n} = (q_m, q_n)$ 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 $\mathcal{Q}_{m,n} = \{\mathbf{q}_{m,n} | 0 \leq q_i \leq P, q_i + q_{\varphi(i)} = P, i = m, n\}$ is the feasible set of transmit power vector for user m and n .

For the simplicity of notation, we let $U(\rho_m)$ stand for the WSR $U(\mathbf{q}_{m,n})$ 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.

Algorithm 2 NOMA-MRT-PA**Input:** $\mu_i, \mathbf{h}_i$ (Assuming $\|\mathbf{h}_m\| \leq \|\mathbf{h}_n\|$), $i = m, n$.**Output:** $\mathbf{p}_{m,n}^*$.

```

1: if  $\|\mathbf{h}_m\| < \|\mathbf{h}_n\|$ 
2:   calculate  $\zeta_{n,m}$  using (8)
3:   if  $\zeta_{n,m} \leq 0$ 
4:      $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,1})\}$ , where  $U(\tilde{\rho}_{n,1})$  is calculated by (41),
        $\tilde{\rho}_{n,1} \in [0, \rho]$ .  $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$ .
5:   else if  $\zeta_{n,m} \geq \rho$ 
6:      $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}$ , where  $U(\tilde{\rho}_{n,2})$  and  $U(\tilde{\rho}_{n,3})$ 
       are calculated by (42),  $\tilde{\rho}_{n,2} \in [0, \rho]$ ,  $\tilde{\rho}_{n,3} \in [0, \rho]$ .  $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$ .
7:   else
8:      $\tilde{\rho}_n^* = \operatorname{argmax}_{\tilde{\rho}_n} \{U(0), U(\rho), U(\zeta_{n,m}), U(\tilde{\rho}_{n,1}), U(\tilde{\rho}_{n,2}), U(\tilde{\rho}_{n,3})\}$ , where
        $\tilde{\rho}_{n,1} \in (\zeta_{n,m}, \rho]$ ,  $\tilde{\rho}_{n,2} \in [0, \zeta_{n,m}]$ ,  $\tilde{\rho}_{n,3} \in [0, \zeta_{n,m}]$ .  $\mathbf{p}_{m,n}^* = [P - \tilde{\rho}_n^* N_0, \tilde{\rho}_n^* N_0]$ .
9:   end if
10: else
11:   if  $\mu_m \geq \mu_n$ ,  $\mathbf{p}_{m,n}^* = [P, 0]$ . Otherwise,  $\mathbf{p}_{m,n}^* = [0, P]$ 
12: end if

```

Theorem 3 The solution of Problem (P3) is the transmit power vector in dual two-user MAC, which corresponds to the maximum among $U(0)$, $U(\rho)$, $U(\rho_{m,1})$, $U(\rho_{m,2})$, $\dots$, $U(\rho_{m,9})$, where $\rho_{m,i} \in [0, \rho]$ 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

$$\begin{aligned}
 U(\rho_m) = & \mu_m \log \left(1 + \frac{\rho_m \|\mathbf{h}_m\|^2 (1 + \rho_n \|\mathbf{h}_n\|^2 \theta)^2}{\zeta_m + \rho_n \|\mathbf{h}_n\|^2 (1 - \theta)} \right) \\
 & + \mu_n \log \left(1 + \frac{\rho_n \|\mathbf{h}_n\|^2 (1 + \rho_m \|\mathbf{h}_m\|^2 \theta)^2}{\zeta_n + \rho_m \|\mathbf{h}_m\|^2 (1 - \theta)} \right),
 \end{aligned} \tag{44}$$

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

$$a_9 \rho_m^9 + a_8 \rho_m^8 + a_7 \rho_m^7 + a_6 \rho_m^6 + a_5 \rho_m^5 + a_4 \rho_m^4 + a_3 \rho_m^3 + a_2 \rho_m^2 + a_1 \rho_m + a_0 = 0 \tag{45}$$

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 from $U(0)$, $U(\rho)$, $U(\rho_{m,1})$, $U(\rho_{m,2})$, $\dots$, $U(\rho_{m,9})$. Consequently, the solution of Problem (P3) is the power allocation vector, which corresponds to the maximum among $U(0)$, $U(\rho)$, $U(\rho_{m,1})$, $U(\rho_{m,2})$, $\dots$, $U(\rho_{m,9})$, where $\rho_{m,i} \in [0, \rho]$. If $\rho_{m,i} \notin [0, \rho]$, $i = 1, 2, \dots, 9$, we let $U(\rho_{m,i}) = 0$. $\square$

By Theorem 3, MMSE-BF-PA is proposed, which is described in Algorithm 3.

Algorithm 3 MMSE-BF-PA**Input:** $\mu_i, \mathbf{h}_i \in \mathbb{C}^{M \times 1}, i = m, n$.**Output:** $\mathbf{p}_{m,n}^*$.

- 1: Calculate $\rho_{m,i}$ using (45), $i = 1, 2, \dots, 9$.
- 2: $\rho_m^* = \operatorname{argmax}_{\rho_m} \{U(0), U(\rho), U(\rho_{m,1}), U(\rho_{m,2}), \dots, U(\rho_{m,9})\}$, where $\rho_{m,i} \in [0, \rho]$.
- 3: $\mathbf{p}_{m,n}^* = [\tilde{p}_m^*, P - \tilde{p}_m^*]$, where $\tilde{p}_m^*$ is calculated by (39), with $q_m^* = \rho_m^* N_0$

5.3 Joint User Pairing and Power Allocation (JUPA) Algorithm

In order to obtain a maximum total sum rate, in the step of forming a user-pair in JUPA, we compare the two sum 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 problem which is how to pair user i with a user from other unpaired users to obtain the maximum sum rate, we can divide all of unpaired users besides user i into 3 subsets, i.e., Φ_1^i, Φ_2^i and Φ_3^i , according to $\zeta_{i,i'}$. Particularly, user i' in Φ_1^i, Φ_2^i and Φ_3^i , satisfies $\zeta_{i,i'} \leq 0, \zeta_{i,i'} \geq \rho$ and $0 < \zeta_{i,i'} < \rho$, respectively. The original user pairing problem is equivalent to 3 problems, which are how to pair user i with a user to obtain the local-maximum sum rate from Φ_1^i, Φ_2^i and Φ_3^i , respectively. Therefore, the solution of the original user pairing problem is the one corresponding to the maximum among the local-maximum sum rates for Φ_1^i, Φ_2^i and Φ_3^i .

2) Considering the problem which is how to pair user i with a user from Φ_1^i (Φ_2^i) 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 Φ_3^i to achieve local-maximum sum rate, we divide Φ_3^i into 2 subsets, i.e., $\Phi_{3,1}^i$ and $\Phi_{3,2}^i$. The user i' in $\Phi_{3,1}^i$ and $\Phi_{3,2}^i$ satisfies $0 < \zeta_{i,i'} \leq \rho/2$ and $\rho/2 < \zeta_{i,i'} < \rho$, respectively. According to (7), the formula $r_{i'} = \log(1 + \frac{\bar{\rho}_{i'} \|\mathbf{h}_i\|^2 (1-\theta)}{1 + \bar{\rho}_i \|\mathbf{h}_i\|^2})$ provides dominant contribution to rate region boundary formed by user i and i' , when $0 < \zeta_{i,i'} \leq \rho/2$. In this case, for the treatability of the problem, we use the formula $r_{i'} = \log(1 + \frac{\bar{\rho}_{i'} \|\mathbf{h}_i\|^2 (1-\theta)}{1 + \bar{\rho}_i \|\mathbf{h}_i\|^2})$ 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 $\Phi_{3,1}^i$, 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 $\Phi_{3,2}^i$ whose channel vector can form the maximum angle with that of user i .

For the simplicity of description of the proposed JUPA in Algorithm 4, assume users are ordered as $\|\mathbf{h}_1\| \leq \dots \leq \|\mathbf{h}_{2K}\|$. Let $\Phi_\kappa^i(k)$ denote the index of k -th entry in Φ_κ^i , $\kappa = 1, 2$. Besides, let $\Phi_{3,\iota}^i(k)$ denote the index of k -th entry in $\Phi_{3,\iota}^i$, $\iota = 1, 2$.

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 Φ_2^i ($\Phi_3^i, \Phi_{3,1}^i$ and $\Phi_{3,2}^i$). **Line 18** is the step

Algorithm 4 Joint User pairing and Power allocation Algorithm (JUPA)

Input: $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$ (assuming $\|\mathbf{h}_1\| \leq \dots \leq \|\mathbf{h}_{2K}\|$), $\mu_i = 1, i = 1, 2, \dots, 2K$.

Output: Ω 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 then
3:     Go to Line 1
4:   end if
5:   for  $j = i + 1 : 2K$  do
6:     If user  $j$  has been paired then
7:       Go to Line 5
8:     end if
9:     If  $\mathbf{h}_i$  and  $\mathbf{h}_j$  is orthogonal then
10:      Obtain  $\mathbf{p}_{i,j}$  by MMSE-BF-PA,  $\Omega = \Omega \cup \{(i, j, \mathbf{p}_{i,j}, 0)\}$ 
11:    end if
12:  end for
13:  Find  $k_1 = \operatorname{argmax}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_1^i(k)})$  and obtain the power allocation solution
     $\mathbf{p}_{i, \Phi_1^i(k_1)}$  for the user-pair  $\{i, \Phi_1^i(k_1)\}$  by NOMA-MRT-PA
14:  Find  $k_2 = \operatorname{argmin}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_2^i(k)})$  and obtain the power allocation solution
     $\mathbf{p}_{i, \Phi_2^i(k_2)}$  for the user-pair  $\{i, \Phi_2^i(k_2)\}$  by NOMA-MRT-PA
15:  Find  $k_{3,1} = \operatorname{argmin}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_{3,1}^i(k)})$  and obtain the power allocation solution
     $\mathbf{p}_{i, \Phi_{3,1}^i(k_{3,1})}$  for the user-pair  $\{i, \Phi_{3,1}^i(k_{3,1})\}$  by NOMA-MRT-PA
16:  Find  $k_{3,2} = \operatorname{argmin}_k \cos^2 \angle(\mathbf{h}_i, \mathbf{h}_{\Phi_{3,2}^i(k)})$  and obtain the power allocation solution
     $\mathbf{p}_{i, \Phi_{3,2}^i(k_{3,2})}$  for the user-pair  $\{i, \Phi_{3,2}^i(k_{3,2})\}$  by NOMA-MRT-PA
17:  Select the power allocation vector with the maximum WSR, from the solutions
    obtained by Line 13-16:

```

$$\mathbf{p}^N = \operatorname{argmax}_{\mathbf{p}} \{U(\mathbf{p}_{i, \Phi_1^i(k_1)}), U(\mathbf{p}_{i, \Phi_2^i(k_2)}), U(\mathbf{p}_{i, \Phi_{3,1}^i(k_{3,1})}), U(\mathbf{p}_{i, \Phi_{3,2}^i(k_{3,2})})\}.$$

Let $\{i, j^N\}$ denote the user-pair which corresponds to $\mathbf{p}^N$.

```

18: Find  $j^M = \operatorname{argmin}_j \cos \angle(\mathbf{h}_i, \mathbf{h}_j)$  and obtain the power allocation solution  $\mathbf{p}_{i, j^M}$ 
    for the user-pair  $\{i, j^M\}$  by MMSE-BF-PA.
19: if  $U(\mathbf{p}^N) \geq U(\mathbf{p}_{i, j^M})$ 
20:    $\Omega = \Omega \cup \{(i, j^N, \mathbf{p}^N, 1)\}$  and  $u_{i, j^N} = u_{j^N, i} = 1$ 
21: else
22:    $\Omega = \Omega \cup \{(i, j^M, \mathbf{p}_{i, j^M}, 0)\}$  and  $u_{i, j^M} = u_{j^M, i} = 1$ 
23: end if
24: end for

```

for similar operation by MMSE-BF-PA. The output of Algorithm 4, Ω , 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 Ω 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.

Note that the computation complexity for obtaining the angle between channels of two users is $\mathcal{O}(M)$, and the number of computing the angle is $2K - 2(i - 1) - 1$ in i -th outer 'for' loop, for MMSE-BF mode and NOMA-MRT mode. Therefore, the overall computation complexity of Algorithm 4 is $\mathcal{O}((\frac{2K}{2})^2 M)$. For exhaustive research based scheme, i.e., Exhaustive Search based user Pairing and Power Allocation (ES-PPA) described in Section VI, the number of all possible pairing schemes is $(2K - 1)!!$. In each pairing scheme, there are K user-pairs and the angle between channels of two users in each user-pair should be calculated. Therefore, the computation complexity of exhaustive research based scheme, is $\mathcal{O}((2K - 1)!! KM)$.

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 is described in algorithm 5, where S is control bit for transmission mode switching and $\Omega(i)_j$ denotes the j -th element in the i -th entry in the set Ω . Noting that TDMA(FDMA) is adopted in Algorithm 5 (Line 10), 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.

Algorithm 5 JUPA/TMS

Input: $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$, $s_i, i = 1, 2, \dots, 2K$.

Output: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$.

- 1: Perform JUPA to obtain pairing and power allocation configuration Ω
 - 2: **for** $i = 1 : K$ **do**
 - 3: $(m, n, \mathbf{p}_{m,n}, S) \leftarrow (\Omega(i)_1, \Omega(i)_2, \Omega(i)_3, \Omega(i)_4)$
 - 4: **if** $S = 1$
 - 5: Calculate $\mathbf{w}_m$ and $\mathbf{w}_n$ with (6)
 - 6: **else if** $S = 0$
 - 7: Calculate $\mathbf{w}_m$ and $\mathbf{w}_n$ with (14)
 - 8: **end if**
 - 9: $\mathbf{x}_i \leftarrow \sqrt{p_m} \mathbf{w}_m s_m + \sqrt{p_n} \mathbf{w}_n s_n$
 - 10: Transmit $\mathbf{x}_i$ 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 (ZF-BF, 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 comparison with traditional transmission schemes in simulation.

6.1 The Adaptive Switching Method

In Fig.2, 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 $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = 1$ and $\alpha = \pi/4$, the rate region of MMSE-BF covers that of NOMA-MRT. From (a) to (d), as MMSE-BF is the optimal linear beamformer in MISO broadcast channel, MMSE-BF get better performance than ZFBF and MRT absolutely, in any channel state.

As demonstrated in Fig.3, 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.

6.2 Joint User Pairing and Power Allocation Based on Transmission Mode Switching

In Fig.4, with different weight vectors, the rate points on rate region boundary and the corresponding maximum-WSR points are obtained by NOMA-MRT-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 point obtained by NOMA-MRT-PA are superposed together with that derived by exhaustive research, which means that NOMA-MRT-PA can derive optimal rate point for different weight vector as shown in (a). In (b), with $\mathbf{u} = (0.5, 0.5)$, the WRS obtained by NOMA-MRT is given when the allocated power of user m varies, with P being normalized to 1. The maximum-WSR point corresponding to the rate point with $\mathbf{u} = (0.5, 0.5)$ in (a) can be obtained by NOMA-MRT-PA, 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.

In Fig.5, with different weight vectors, rate points on rate region boundary and corresponding maximum-WSR points are obtained by MMSE-BF-PA and exhaustive search, respectively. Similarly, with different weight vectors, the rate points obtained by MMSE-BF-PA are superposed with those derived by exhaustive research, and MMSE-BF-PA can derive optimal rate point for different weight vector as shown in (a). In (b), with $\mathbf{u} = (0.5, 0.5)$, the WRS obtained by MMSE-BF is given when the allocated power of user m varies, and the maximum-WSR point can also be obtained by MMSE-BF-PA, which is 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.

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 [24], [25] 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 $N_0 = -10\text{dBm}$, the cell radius $R = 30\text{m}$, 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

$$j^* = \operatorname{argmax}_j \max\{U_{i,j}^{\text{NOMA-MRT}}, U_{i,j}^{\text{MMSE-BF}}\}, j \neq i$$

where $U_{i,j}^{\text{NOMA-MRT}}$ and $U_{i,j}^{\text{MMSE-BF}}$ stand for the sum rate of user i and j , obtained by NOMA-MRT-PA and MMSE-BF-PA, respectively. 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

$$j^* = \operatorname{argmax}_j \frac{|\mathbf{h}_j^H \mathbf{h}_i|^2}{\|\mathbf{h}_i\|^2 \|\mathbf{h}_j\|^2}, j \neq i.$$

Similarly, 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 the sum rates of K user-pairs.

In Fig.6 and 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 antenna at the S varies, respectively. In Fig.6, the sum rates of all schemes grow with the transmit power P increasing. Compared with optimal ES-PPA using exhaustive search to obtain the user pairing solution, the proposed JUPA only suffers very slight performance loss. However, the computation complexity of proposed JUPA is only $\mathcal{O}(K^2 M)$, while ES-PPA is $\mathcal{O}((2K - 1)!! KM)$. In comparison with other schemes, the JUPA derives better performance and the performance gain boosts with P/N_0 increasing. Further, JUPA outperforms GA-PPA, since GA dose 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 Fig.7, it is can be observed that the proposed JUPA results in a slight performance loss compared with optimal ES-PPA. However, the JUPA outperforms other schemes and the performance gain over other schemes decreases as the number of antenna at the BS increases.

7 Conclusion

In this paper, the design of adaptive TMS between NOMA-MRT and MMSE-BF to maximize the sum rate for MISO-NOMA 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, by exploiting the optimal solutions by the aforementioned power allocation algorithms for two transmission modes and the high efficiency of TMS, 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 computation complexity of $\mathcal{O}((2K-1)!!KM)$, the proposed JUPA can obtain a much lower complexity of $\mathcal{O}(K^2M)$ 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

$$\begin{cases} r_n = \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2) \\ r_m = \min\{\log(1 + \gamma_{m,n}^B), \log(1 + \gamma_m^B)\}, \end{cases} \quad (46)$$

$$\quad (47)$$

where

$$\begin{cases} \gamma_{m,n}^B = \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2} \\ \gamma_m^B = \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)}, \end{cases} \quad (48)$$

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

Defines

$$f(\tilde{\rho}_n) = \gamma_{m,n}^B - \gamma_m^B. \quad (49)$$

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

$$m_1 \tilde{\rho}_n^2 + m_2 \tilde{\rho}_n + m_3 = 0 \quad (50)$$

where $m_1 = \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2)$, $m_2 = -(\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2) - \rho \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^2 (2\theta - \theta^2)$, $m_3 = \rho (\|\mathbf{h}_n\|^2 (1 - \theta) - \|\mathbf{h}_m\|^2)$.

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

$$\begin{aligned}\Delta &= m_2^2 - 4m_1m_3 \\ &= (\rho\|\mathbf{h}_n\|^2\|\mathbf{h}_m\|^2(2\theta - \theta^2) - (\|\mathbf{h}_n\|^2(1 - \theta) - \|\mathbf{h}_m\|^2))^2.\end{aligned}$$

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

$$\begin{aligned}\tilde{\rho}_{n,1}^{(\Delta)} &= \rho \\ \tilde{\rho}_{n,2}^{(\Delta)} &= \zeta_{n,m}\end{aligned}\tag{51}$$

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

As $m_1 > 0$, 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 $\log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in [0, \rho]$. In this case, $r_m = \log(1 + \gamma_m^B)$ 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. $\tilde{\rho}_{n,2}^{(\Delta)} \leq 0$, the function $f(\tilde{\rho}_n) \leq 0$ with $\tilde{\rho}_n \in [0, \rho]$, which means that $\log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_m^B)$ 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 $\log(1 + \gamma_{m,n}^B) \geq \log(1 + \gamma_m^B)$ with $\tilde{\rho}_n \in [0, \zeta_{n,m}]$ and $\log(1 + \gamma_{m,n}^B) \leq \log(1 + \gamma_m^B)$ 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

$$\begin{cases} \gamma_{m,n}^B = \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2} \\ \gamma_m^B = \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2}. \end{cases}$$

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

Proof of Lemma 2

The upper bound $r_i^{\text{upper}} = \log(1 + Pl^2/N_0)$ 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 $(r_m^{\text{upper}}, 0), (0, r_n^{\text{upper}})$. As a result, when $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, by time-sharing between the two extreme rate tuple, the achievable rate region boundary of MISO broadcast channel can be expressed as [22]

$$\begin{cases} r_n^{\text{TS}} = \lambda \log\left(1 + \frac{Pl^2}{N_0}\right) \\ r_m^{\text{TS}} = (1 - \lambda) \log\left(1 + \frac{Pl^2}{N_0}\right), \end{cases}\tag{52}$$

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 $\|\mathbf{h}_m\| = \|\mathbf{h}_n\| = l$, the achievable rate region boundary can be characterized as

$$\begin{cases} r_n = \log\left(1 + \frac{p_n l^2}{N_0}\right) \\ r_m = \log\left(1 + \frac{p_m l^2 \cos^2 \alpha}{N_0 + p_n l^2}\right). \end{cases} \quad (53)$$

$\forall l \in (0, \infty), \alpha \in [0, \pi), \lambda \in [0, 1], p_m \in [0, P]$ and $p_n \in [0, P]$, we let

$$r_n^{\text{TS}} = r_n. \quad (54)$$

We define

$$f(\lambda, p_m) = r_m^{\text{TS}} - r_m. \quad (55)$$

By substituting (54) into (55), we can get the result

$$f(\lambda, p_m) = \frac{p_m l^2 (1 - \cos^2 \alpha)}{N_0 + (1 - p_m) l^2} \geq 0. \quad (56)$$

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 $\zeta_{n,m} \leq 0$, the WSR $U(\tilde{\rho}_n)$ can be expressed as

$$U(\tilde{\rho}_n) = \mu_n \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2) + \mu_m \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_n\|^2 (1 - \theta)}{1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2}\right), \quad (57)$$

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

$$\varepsilon_1 \tilde{\rho}_n + \varepsilon_0 = 0$$

where $\varepsilon_1 = \mu_n \|\mathbf{h}_n\|^2 \theta$, $\varepsilon_0 = (1 + \rho \|\mathbf{h}_n\|^2)(\mu_n - \mu_m(1 - \theta)) - \mu_n \|\mathbf{h}_n\|^2 \rho \theta$. As a result, we can obtain (41). In this case, the global maximum value of $U(\tilde{\rho}_n)$ in the interval $[0, \rho]$ can be found by selecting the maximum from $U(0)$, $U(\rho)$ and $U(\tilde{\rho}_{n,1})$. Therefore, the solution of Problem (P2) is the power allocation vector corresponding to the maximum among $U(0)$, $U(\rho)$ and $(\tilde{\rho}_{n,1})$.

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

$$U(\tilde{\rho}_n) = \mu_n \log(1 + \tilde{\rho}_n \|\mathbf{h}_n\|^2) + \mu_m \log\left(1 + \frac{\tilde{\rho}_m \|\mathbf{h}_m\|^2}{1 + \tilde{\rho}_n \|\mathbf{h}_m\|^2 (1 - \theta)}\right), \quad (58)$$

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

$$\epsilon_2 \tilde{\rho}_n^2 + \epsilon_1 \tilde{\rho}_n + \epsilon_0 = 0 \quad (59)$$

where $\epsilon_2 = \mu_n \|\mathbf{h}_n\|^2 \|\mathbf{h}_m\|^4 \theta (1 - \theta)$, $\epsilon_1 = \|\mathbf{h}_n\|^2 (\mu_m \epsilon_3 - \mu_n \epsilon_4)$, $\epsilon_0 = \mu_m \epsilon_3 - \mu_n \|\mathbf{h}_n\|^2 (1 + \|\mathbf{h}_m\|^2 \rho)$, $\epsilon_3 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) + 1)$ and $\epsilon_4 = \|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \rho (1 - \theta) - 2\theta + 1)$. By using the quadratic formula to solve Eq. (59), we can obtain (42). Consequently, the solution of Problem (P2) is the power allocation vector, which corresponds to the maximum among $U(0)$, $U(\rho)$, $U(\tilde{\rho}_{n,2})$ and $U(\tilde{\rho}_{n,3})$.

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 the power allocation vector corresponding to the maximum among $U(0)$, $U(\rho)$, $U(\zeta_{n,m})$, $U(\tilde{\rho}_{n,1})$, $U(\tilde{\rho}_{n,2})$ and $U(\tilde{\rho}_{n,3})$, where $\rho_{n,1} \in (\zeta_{n,m}, \rho]$, $\rho_{n,i} \in [0, \zeta_{n,m}]$, $i = 2, 3$.

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, $r_m + r_n = \log(1 + \rho l^2)$. Consequently, if $\mu_m \geq \mu_n$, $\mu_m r_m + \mu_n r_n \leq \mu_m \log(1 + \rho l^2)$, i.e., $U(0) = \mu_m \log(1 + \rho l^2)$ is the global maximum value of $U(\tilde{\rho}_n)$ in the interval $[0, \rho]$, implying that the solution of Problem (P2) is $\mathbf{p}_{m,n}^* = [P, 0]$.

If $\mu_m < \mu_n$, the proof is similar to that of the above case, in which $\mu_m \geq \mu_n$.

Abbreviations

TMS:transmission mode switching; MISO:multiple-input-single-output;NOMA:non-orthogonal multiple access; NOMA-MRT:NOMA-based maximum ratio transmission; BF:beamforming; MMSE-BF:minimum mean square error beamforming; ZFBF:zero-forcing beamforming; WSR:weighted sum rate; MAC:multiple-access channel; CSI:the channel state information; BC:broadcast channel; SIC:succesive interference cancellation; NOMA-MRT-PA:Power Allocation algorithm for NOMA-MRT mode; MMSE-BF-PA:Power Allocation algorithm for MMSE-BF mode; BS:base station; SINR:signal-to-interference-plus-noise ratio; JUPA:Joint User pairing and Power Allocation algorithm; ES-PPA:Exhaustive Search based user Pairing and Power Allocation algorithm; GA-PPA:Greedy Algorithm based user Pairing and Power Allocation algorithm; COR-PPA:CORrelation-based-pairing based user Pairing and Power Allocation algorithm; RAN-PPA:RANDOM pairing based user Pairing and Power Allocation algorithm; AWGN:additive white Gaussian noise.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Authors' contributions

All authors contributed to the design and analysis of the research, to the simulation results, and to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

This work is supported in part by the Non-funded science and technology project of Zhanjiang under Grant 2020B01067; in part by the Special Foundation of Talents Project at Lingnan Normal University under Grant ZL2048; in part by the Fundamental Research Funds project for provincial universities in Heilongjiang Province under Grant Hkdxp201804.

Availability of data and materials

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

Author details

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

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., Karagiannis, 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., Karagiannis, G.K.: On the application of quasi-degradation to MISO-NOMA downlink. *IEEE Transactions on Signal Processing* **64**(23), 6174–6189 (2016)
15. 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
16. 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
17. 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)
18. 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)
19. 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)
20. Ding, Z., Liu, Y., Choi, J., Sun, Q., El Kashlan, 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)
21. Sklar, B.: Rayleigh fading channels in mobile digital communication systems. I. characterization. *IEEE Communications magazine* **35**(7), 90–100 (1997)
22. Tse, D., Viswanath, P.: *Fundamentals of Wireless Communication*. Cambridge university press, Cambridge (2005)
23. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge university press, Cambridge (2012)
24. Kimy, B., Lim, S., Kim, H., Suh, S., Kwun, J., Choi, S., Lee, C., Lee, S., Hong, D.: Non-orthogonal multiple access in a downlink multiuser beamforming system. In: MILCOM 2013-2013 IEEE Military Communications Conference, pp. 1278–1283 (2013). IEEE
25. Choi, J.: Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems. *IEEE Transactions on Communications* **63**(3), 791–800 (2015)

Figures

Figure 3 Rate region of the Adaptive Switching Method with time-sharing.

$P/N_0 = 10\text{dB}$, $\|\mathbf{h}_m\| = 1$, $\|\mathbf{h}_n\| = 2$, $\alpha = \pi/4$. Also plotted for comparison are rate regions of ZFBF, MRT, MMSE-BF and NOMA-MRT, without time-sharing.

Figure 4 Rate points and the corresponding maximum-WSR points obtained by NOMA-MRT-PA and exhaustive research. $P/N_0 = 0\text{dB}$, $\|\mathbf{h}_m\| = 2$, $\|\mathbf{h}_n\| = 1$, $\alpha = \pi/4$. (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.

Figure 5 Rate points and corresponding maximum-WSR points obtained by MMSE-BF-PA and exhaustive search. $P/N_0 = 10\text{dB}$, $\|\mathbf{h}_m\| = 2$, $\|\mathbf{h}_n\| = 1$, $\alpha = \pi/4$. (a) rate points under different weight vectors. (b), (c) and (d) are for maximum-WSR points with $\mathbf{u} = (0.5, 0.5)$, $\mathbf{u} = (0.33, 0.67)$ and $\mathbf{u} = (0.25, 0.75)$.

Figure 6 Sum rate versus P for different transmission schemes. $M=2$, $2K=8$.

Figure 7 Sum rate versus number of antenna for several transmission schemes. $P=40\text{dBm}$, $2K=8$.