Liang *et al.*
EURASIP Journal on Advances in Signal Processing (2024) 2024:1
https://doi.org/10.1186/s13634-023-01099-8

EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Neighbor-based joint spatial division and multiplexing in massive MIMO: user scheduling and dynamic beam allocation

Huibin Liang[1], Chen Liu[1*], Yunchao Song[1], Tianbao Gao[1] and Yulong Zou[2]

*Correspondence: liuch@njupt.edu.cn

[1] College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, No. 9 Wenyuan Road, Qixia District, Nanjing 210003, People's Republic of China
[2] School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China

Abstract

Two-stage precoding schemes have been developed to reduce the channel estimation overhead in FDD systems. By integrating user scheduling into these schemes, it becomes possible to meet the quality-of-service requirements of high-density wireless communication systems, despite the limitations on spatial resources and transmit power budget. In this paper, we propose a user scheduling and dynamic beam allocation method for neighbor-based joint spatial division multiplexing (N-JSDM) transmission. The user scheduling problem is formulated as a 0–1 quadratic programming problem to maximize effective spectral efficiency (ESE) using directional channel properties. To address the complexity issue, convex relaxation and linearization methods are employed to transform the problem into a 0–1 mixed integer linear programming, and a dimensionality reduction method is introduced. The proposed user scheduling-aided N-JSDM scheme reduces downlink training length and feedback of channel state information. Additionally, a dynamic configuration form is used for pre-beamforming matrix design based on user distribution, outperforming conventional approaches. Simulation results demonstrate higher ESE achieved by the proposed scheme compared to previous methods.

Keywords: N-JSDM, User scheduling, Dynamic beam allocation, Mixed integer programming, Linearization

1 Introduction

Over the past three decades, the data rates of wireless communication have been doubling every eighteen months, and it is projected to reach Terabit-per-second in the near future [1]. Massive multiple-input multiple-output (MIMO) has been a crucial technology for enhancing system throughput and providing reliable communication [2]. By employing a large-scale antenna array at the base station (BS), massive MIMO achieves higher data transmission rates, with the number of BS antennas significantly surpassing the number of served user terminals. It utilizes spatial resources and capitalizes on the multipath propagation characteristics to establish a parallel transmission mechanism, multiplying system capacity without the need for additional spectrum resources

or transmit power [3]. In the forthcoming communication systems, massive MIMO will continue to play a pivotal role.

Massive MIMO relies on the channel state information (CSI), which is the communication link state information from the transmitter to the receiver [4]. When the CSI is perfect, the performance of massive MIMO scales linearly with the smaller number of antennas between the transmit and receive sides [5], underscoring the critical importance of obtaining instantaneous CSI. In previous research on massive MIMO systems, time division duplex (TDD) mode has been widely adopted. TDD leverages channel reciprocity, enabling the estimation of downlink CSI through the uplink channel, thereby reducing spectral overhead [6–9]. However, the prevailing wireless standards predominantly employ frequency division duplex (FDD) systems, which offer more mature industrial products and market share [10]. Furthermore, in the extensively studied millimeter-wave frequency band, FDD systems may exhibit similarly impressive performance in cell-free massive MIMO systems [11]. Nonetheless, due to the absence of channel reciprocity, FDD massive MIMO systems necessitate substantial downlink training length (DTL) and CSI feedback during the downlink communication to acquire CSI at the transmitter [12]. Additionally, the cost of reconfiguring frequency bands to accommodate TDD in FDD systems is considerably high [13]. Consequently, for FDD massive MIMO systems, acquiring CSI presents a significant challenge, particularly for telecom operators compelled to upgrade their existing FDD systems to 5 G wireless communications.

There have been a lot of research efforts on reducing DTL and channel feedback in FDD massive MIMO systems. Similar to the CSI acquisition in TDD mode, several studies (e.g., [11] and [13]) leverage angle reciprocity by transmitting uplink pilots to obtain CSI, thereby eliminating the need for CSI feedback. The minimum number of pilots required corresponds to the number of terminals. Moreover, some works focus on the spatially correlated MIMO channels and utilize the structure of CSI to reduce DTL and CSI feedback. Specifically, the compressed sensing techniques are used to exploit channel sparsity [14, 15]. Expanding on the consideration of spatial correlation, additional researches have taken into account temporal correlation and leveraged the spatial and temporal common sparsity of massive MIMO channels to acquire CSI with reduced overhead [4, 16]. Additionally, a two-stage beamforming scheme called joint spatial division multiplexing (JSDM) based on statistical CSI is proposed [12]. The JSDM beamforming scheme comprises two stages. In the first stage, the pre-beamformer uses the channel covariance matrix (CCM) to mitigate inter-group interference. In the second stage, the instantaneous CSI of each group is used to design a precoding scheme for intra-group interference suppression. Obtaining the statistical CSI is relatively easier compared to instantaneous CSI since its variations occur at a slower rate [17, 18].

Extensive research attentions have been paid to enhance the performance of JSDM [19–23]. Some works consider the pre-beamformer design to achieve a better spectral efficiency [19–21]. Specifically, due to the non-convexity caused by signal-to-interference-plus-noise ratio (SINR) as an optimization criterion, Kim et al. proposed to use signal-to-leakage-plus-noise ratio (SLNR) as the optimization objective and simplified the SLNR-based pre-beamformer design problem to the trace quotient problem encountered in the field of machine learning [19]. In [20], Jeon et al. used the

minimum mean squared error criterion to design the pre-beamformer and multi-user precoder sequentially. However, none of the above works considered the impact of user grouping. Since the channel covariance matrices of users differ, and the goal of user grouping is to make users in each group have a common eigen-subspace, there will inevitably be overlapping signal spaces between groups. Eliminating inter-group interference by pre-beamforming will reduce the signal space and result in a loss of system performance. Recently, a scheme called neighbor-based JSDM (N-JSDM) is proposed in [21], which avoids the user grouping problem by adopting the neighbor scheme to fully utilize the signal space. N-JSDM is still a two-stage scheme. In the first stage, a pre-beamforming matrix is designed according to the CCMs to reduce the interference of non-neighbors, and the effective channel matrix becomes a band matrix. Neighbor interference is removed in the second stage. Besides, Khalilsarai et al. proposed a method to approximate the downlink CCM of users as the columns of the discrete Fourier transformation matrix, particularly when the number of antennas at the BS is large [22]. This approximation enables the BS to utilize codebook-based beam selection for designing the pre-beamforming matrix, thereby reducing the computational complexity. There are also works to improve the performance of JSDM from the aspects of antenna structures [23] and BS selection [24]. Tang et al. provided an analysis of two-stage precoding designs under different antenna structures, offering guidelines for antenna structure selection to achieve a better balance between performance and cost [23]. Considering that the overlap of the angle-spreading-ranges (ASR) of different user clusters may seriously degrade the performance of two-stage precoding, Ma et al. proposed a solution to minimize ASR using BS selection [24].

As the number of users increases in the system, inter-user interference becomes severe, and a portion of degrees of freedom is used to mitigate inter-group interference, resulting in a degradation of desired signal energy [21]. Therefore, it is necessary to schedule users to improve the spectral efficiency. User scheduling in conventional JSDM schemes are divided into two parts: user grouping and intra-group user scheduling. Before implementing JSDM beamforming, users need to be grouped, and the users in each group share a common eigen-subspace, i.e., group eigen-space, where the group eigen-spaces of different groups are orthogonal or non-overlapping. Several user grouping methods have been proposed [25–28]. For example, the K-means clustering algorithm based on chordal distance and fixed quantization algorithm based on discrete Fourier transform are proposed in [25]. Xu et al. presented a K-means algorithm based on weighted likelihood metric in [26]. Nam et al. transformed user grouping into a subspace packing problem in Grassmann manifold [27], while a recent work [28] proposes a hierarchical clustering algorithm that considers both the number of groups and the chordal distance threshold. Besides, intra-group user scheduling has also been studied. A scheduling algorithm based on average SLNR has been proposed [28]. The iterative SLNR-based group scheduling combines the outer precoder and group scheduling to achieve better performance. Xu et al. proposed an optimized scheduling algorithm based on channel quality indicator (CQI). The algorithm assumes that the users cannot achieve the maximum value on two or more beams and assigns a specific beam to each user based on CQI, allowing the user to obtain maximum gain on that beam [26].

Considering the advantages of N-JSDM, incorporating user scheduling into the N-JSDM transmission scheme enables better integration of precoding techniques, further optimizing system performance and enhancing communication quality. In this paper, we propose a user scheduling and dynamic beam allocation method for the N-JSDM transmission scheme to maximize effective spectral efficiency (ESE) subject to limited pilot length. Specifically, considering the challenges in acquiring complete CSI, we formulate the user scheduling problem as a 0–1 quadratic programming by leveraging the channel directional features. Since the users are randomly distributed, we propose dynamically allocating the number of beams serving each user. This idea is incorporated into the optimization problem as a constraint, and the pre-beamformer is designed accordingly. Additionally, we transform the optimization problem into a 0–1 mixed integer programming problem using convex relaxation and linearization techniques. Simulation results demonstrate the validity of the theoretical analysis. The primary contributions of this paper can be summarized as follows:

- We analyze the factors that impact ESE and formulate the user scheduling problem as a 0–1 quadratic programming problem with linear constraints, leveraging the channel directional features. These features are more stable over larger time scales compared to instantaneous CSI, which varies according to the channel coherence time.
- To simplify the 0–1 quadratic programming problem, we employ convex relaxation and linearization techniques to transform it into a mixed integer linear programming problem. Additionally, to further reduce computational complexity, we propose a dimensionality reduction method.
- The pre-beamformer design using dynamic allocation scheme is proposed. Since the number of beams serving each user is determined by the interference between the user and its neighbors, it can be well applied in realistic scenarios where users are randomly distributed and/or DTL is limited.

The rest of the paper is organized as follows. Section 2 describes the system model and the N-JSDM scheme. The problem formulation of N-JSDM user scheduling is provided in Sect. 3. The beam allocation method based on overlap density, the linearization method of 0–1 quadratic programming, and the dimensionality reduction method are presented in Sect. 4. In Sect. 5, we propose a pre-beamformer design with dynamic beam configuration. Simulation results and discussion are given in Sect. 6. Finally, we conclude this paper in Sect. 7.

1 Notation

Bold uppercase letters indicate matrices, and bold lowercase letters represent column vectors. The i-th row and i-th column of matrix $\mathbf{A}$ *are denoted by* $\mathbf{a}^i$ *and* $\mathbf{a}_i$*, respectively. The factorial of a is represented by* $a!$*.* $\mathbf{I}_N$ *represents the* $N \times N$ *identity matrix, while the superscripts* $(\cdot)^H$ *and* $(\cdot)^T$ *denote the conjugate transpose and transpose of a matrix, respectively. The pseudo-inverse operation is denoted by* $(\cdot)^\dagger$*. The orthogonal complement*

space is represented by span$^{\perp}(\cdot)$*. The Hadamard product is denoted by* $\odot$*. The set of real numbers and complex numbers is* $\mathbb{N}$ *and* $\mathbb{C}$*, respectively.* ι *represents the imaginary unit, i.e.,* $\iota = \sqrt{-1}$*.*

2 Preliminary

2.1 System model

We consider a typical single-cell FDD massive MIMO system where a BS is equipped with a uniform linear array (ULA) of M elements serving K single-antenna users. The BS applies a precoder $\mathbf{V} \in \mathbb{C}^{M\times K}$ in the downlink to transmit symbols. Then, the received signal at the users can be written as

$$\mathbf{y} = \mathbf{H}^H\mathbf{V}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T \in \mathbb{C}^{K\times 1}$ with y_k being the received signal of user k, $\mathbf{H}^H = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]^H \in \mathbb{C}^{K\times M}$ is the channel matrix with $\mathbf{h}_k^H \in \mathbb{C}^{1\times M}$ being the channel from BS to user k, $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T \in \mathbb{C}^{K\times 1}$ is the transmitted signal satisfying a power constraint $E(\mathbf{s}\mathbf{s}^H) = \mathbf{I}_K$, and $\mathbf{n} = [n_1, n_2, \ldots, n_K] \in \mathbb{C}^{K\times 1}$ denotes the additive white Gaussian noise vector with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$.

In this paper, we adopt a one-ring (OR) channel model [29], in which user k has an azimuth angle θ and an angular spread (AS) Δ, and θ_k is randomly distributed[1]. Here, we sort the users as $\theta_1 \le \theta_2 \le \cdots \le \theta_K$. The (m,p)-th element of the CCM $\mathbf{C}_k$ of user k is [29]

$$[\mathbf{C}_k]_{m,p} = \frac{1}{2\Delta}\int_{\theta_k-\Delta}^{\theta_k+\Delta} e^{\frac{-\iota 2\pi D(m-p)\sin\theta}{\lambda_C}}\,\mathrm{d}\theta, \tag{2}$$

where λ_C is the carrier wavelength, $D = \lambda_C/2$ is the spacing between two antenna elements. According to Karhunen–Loeve representation, we can write the channel vector of user k as $\mathbf{h}_k = \mathbf{C}_k^{1/2}\mathbf{z}_k$, where $\mathbf{z}_k$ is small-scale fading with $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Letting $\mathbf{C} = \sum_{k=1}^{K}\mathbf{C}_k$, $span(\mathbf{H})$ can be any subspace of $span(\mathbf{C})$.

Since statistical CSI varies much slower than the instantaneous CSI, the BS can accurately obtain the statistical CSI through long-term feedback [32, 33].

2.2 The description of N-JSDM scheme

The N-JSDM uses neighbor scheme instead of grouping scheme to fully utilize the signal space and thus provide a better performance. The following is a brief introduction about N-JSDM.

For user k, if $|\theta_k - \theta_j| > \omega$, then user j is called user k's non-neighbor, and the index set of user k's non-neighbors is $\bar{\Omega}_k = \{j| \mid \theta_k - \theta_j \mid > \omega\}$, where ω is called neighbor angular spread (NAS) (in [21], ω is chosen to be 2Δ); the index set of user k and its neighbors is $\Omega_k = \{j| \mid \theta_k - \theta_j \mid \le \omega\}$. Since $\theta_1 \le \theta_2 \le \cdots \le \theta_K$, the elements in Ω_k are consecutive numbers, and Ω_k can be represented as $\Omega_k = \{k_l, \ldots, k, \ldots, k_u\}$. In the following, we refer to the set Ω_k as the neighbor domain of user k.

[1] We assume that the information about the CCM is known and accurate. This assumption is reasonable because there have been research studies on CCM estimation, as detailed in [30, 31]. These works leverage the angular reciprocity between the uplink and downlink channels in FDD systems to improve channel estimation.

N-JSDM is a two-stage beamforming scheme. In the first stage, user k's CCM $\mathbf{C}_k (k = 1, 2, \ldots, K)$ is used to design the pre-beamforming matrix $\mathbf{B}_k$ to reduce non-neighbor interference, so that for each k

$$\mathbf{h}_k^H \mathbf{B}_i = \mathbf{0}, i \in \bar{\Omega}_k. \tag{3}$$

The effective channel matrix after the pre-beamforming stage is $\mathbf{H}^H \mathbf{B}$ where $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K]$. From Eq. (3), the k-th row of $\mathbf{H}^H \mathbf{B}$ can be written as

$$\begin{aligned} \mathbf{h}_k^H \mathbf{B} &= (\mathbf{h}_k^H \mathbf{B}_1, \mathbf{h}_k^H \mathbf{B}_2, \ldots, \mathbf{h}_k^H \mathbf{B}_K) \\ &= (\cdots, 0, 0, \mathbf{h}_k^H \mathbf{B}_{\Omega_k}, 0, 0, \ldots), \end{aligned} \tag{4}$$

where $\mathbf{B}_{\Omega_k} = [\mathbf{B}_{k_l}, \ldots, \mathbf{B}_k, \ldots, \mathbf{B}_{k_u}]$ is defined as the matrix composed of the pre-beamforming matrix of user k and its neighbors. Equation (4) indicates that $\mathbf{h}_k^H \mathbf{B}$ has a continuous sequence of $col(\mathbf{B}_{\Omega_k})$ nonzero values, where $col(\cdot)$ refers to the number of columns. It should be noted that since the azimuth angle of users is sorted, when $\theta_k > \theta_j$, there must be $k_l \geq j_l$ and $k_u \geq j_u$, so the effective channel matrix $\mathbf{H}^H \mathbf{B}$ is a band matrix.

In the second stage of N-JSDM, to eliminate the interference from neighbors, $\mathbf{W} = (\tilde{\mathbf{H}})^\dagger \mathbf{\Gamma}$ is designed using the zero forcing criterion [34]. Here, $\tilde{\mathbf{H}}$ represents the estimation of the effective channel matrix $\mathbf{H}^H \mathbf{B}$, and $\mathbf{\Gamma} = diag(\gamma_1, \gamma_2, \ldots, \gamma_K)$ is used to normalize each column of $\tilde{\mathbf{H}}$. As a result, the SINR at user k is given by [21]

$$SINR_k = \frac{| \mathbf{h}_k^H \mathbf{B} \mathbf{w}_k |^2}{\sum_{k' \neq k} | \mathbf{h}_k^H \mathbf{B} \mathbf{w}_{k'} |^2 + \sigma^2}, \tag{5}$$

where $\mathbf{w}_k$ is the k-th column of $\mathbf{W}$ and σ^2 is the noise power. It should be noted that the precoding matrix of N-JSDM is written as $\mathbf{V} = \mathbf{B}\mathbf{W}$.

In more practical scenarios, users are randomly distributed. The conventional JSDMs divide users into G groups, and when the users are distributed randomly, there always exists common space between the signal spaces of adjacent groups. The signal space of the g-th group is denoted by $span(\mathbf{H}_g)$. To mitigate the inter-group interference, $span(\mathbf{B}_g)$ is orthogonal to all the signal space $span(\mathbf{H}_j)$, $j \neq g$. This means that $span(\mathbf{B}_g)$ is orthogonal to all the overlapped signal space, and hence $\cup_{g=1,2,\cdots,G} span(\mathbf{B}_g)$ (i.e., $span(\mathbf{B})$) is orthogonal to all the overlapped signal space. Consequently, $span(\mathbf{H}) \nsubseteq span(\mathbf{B})$, resulting in a lower-dimensional utilized signal space $span(\mathbf{B}^H \mathbf{H})$ compared to the full signal space $span(\mathbf{H})$, thereby decreasing the performance of JSDM. Compared to conventional JSDMs, N-JSDM offers the following advantages: Firstly, it achieves higher spectral efficiency. N-JSDM employs a neighbor grouping approach to further divide users into subgroups, eliminating the requirement for users in the same neighbor domain to share the same common subspace. This allows for the use of more refined precoding techniques to reduce interference, thereby improving the system's spectral efficiency. Secondly, N-JSDM exhibits better interference mitigation capabilities. When designing the pre-beamforming scheme, N-JSDM takes into account the mutual interference between subgroups. By optimizing the pre-beamforming matrix, interference between subgroups can be more effectively reduced, enhancing the system's interference mitigation performance. Therefore, N-JSDM is considered a more promising and feasible beamforming scheme.

3 User scheduling in N-JSDM

The N-JSDM with user scheduling involves three stages. The first stage is user scheduling, which is used to determine the azimuth angle of the scheduled users and the number of beams serving each user, denoted by θ_u and g_u, respectively. The second and third stages are pre-beamforming and multi-user precoding, which are used to obtain matrix $\mathbf{B}$ and matrix $\mathbf{W}$, respectively.

In a more realistic scenario with only a limited number of users are randomly distributed in the cell, it is more reasonable to dynamically allocate the number of beams serving each user. Therefore, in the user scheduling stage, we propose to allocate beams according to the interference between users and their neighbors. The idea of dynamic allocation is also extended to the pre-beamforming stage. In the pre-beamforming stage, the obtained θ_u and g_u about the scheduled users are used to design the pre-beamformer. Further details regarding the design of the pre-beamformer can be found in Sect. 5. The transformation process of addressing the user scheduling problem is outlined below.

Because there is no concept of user group, the user scheduling approach in N-JSDM is fundamentally different from that of conventional JSDM. To address this, we propose a user scheduling algorithm that solely relies on user directional features. Specifically, we use two channel directional features [35]: azimuth angle and AS. Compared to the instantaneous CSI, these features are more stable [32, 33], and easier to obtain.

3.1 Problem formulation

The objective of scheduling is to maximize the ESE of the system. Assuming a coherence block with T_C symbols and pilot length P, the ESE of user k can be expressed as

$$R_k = \left(1 - \frac{P}{T_C}\right) \log_2(1 + \mathrm{SINR_k}). \tag{6}$$

Note that there is overlapping signal space between some users in the system, and such overlapping signal space represents the inter-user interference (IUI). In the OR channel model, the angle region of user k is defined as

$$\Phi_k = (\theta_k - \Delta, \theta_k + \Delta). \tag{7}$$

Based on the angle region, we introduce the overlap angle (OA) to represent the degree of overlap between users. If there is an intersection between the angle regions of user k and user j, i.e., $\Phi_k \cap \Phi_j \neq \emptyset$, the intersection is called an OA. The OA of user k and user j is defined as $(j, k = 1, 2, \ldots, K)$

$$A_{kj} = \begin{cases} -\mid \theta_k - \theta_j \mid + 2\Delta, & \mid \theta_k - \theta_j \mid \leq 2\Delta, \mathrm{j} \neq \mathrm{k}; \\ 0, & else. \end{cases} \tag{8}$$

These angles depict the interference among users. Since $\omega = 2\Delta$, the OA between user k and user j is nonzero if they are neighbors, and zero otherwise. By using the OA, we can construct an angle matrix $\mathbf{A}$, where A_{kj} is the (k, j)-th element of matrix $\mathbf{A}$. The k-th row of the matrix $\mathbf{A}$ can be written as

$$\mathbf{A}^k = (A_{k1}, A_{k2}, \ldots, A_{kK}) = (\cdots, 0, 0, \mathbf{A}_{k\Omega_k}, 0, 0, \ldots), \tag{9}$$

where $\mathbf{A}_{k\Omega_k} = [A_{kk_l}, \ldots, A_{kk_u}]$ is composed of the OA of user k and its neighbors. From (9), it can be observed that $\mathbf{A}_{k\Omega_k}$ has $|\Omega_k| - 1$ nonzero elements, where $|\Omega_k|$ refers to the number of elements in the index set Ω_k. To distinguish the neighbors and non-neighbors, we introduce an unweighted matrix $\hat{\mathbf{A}}$, whose k-th row can be written as

$$\hat{\mathbf{A}}^k = (0, \ldots, 0, \underbrace{1, 1, \ldots, 1, \overbrace{0}^{\text{user}k}, 1, \ldots, 1}_{|\Omega_k|-1}, 0, \ldots, 0). \tag{10}$$

Since the denominator of SINR contains the interference term, there is a strong correlation between IUI and SINR. By reducing interference among users, SINR increases. Furthermore, in practical systems, the length of pilot sequences is often limited. As a result, the problem of maximizing the ESE is transformed into minimizing interference while adhering to the constraint of maximum pilot length.

To describe the problem, use x_i to denote whether user i is selected, i.e.,

$$x_i = \begin{cases} 1, & \textit{selected}; \\ 0, & \textit{not selected}. \end{cases} \tag{11}$$

Then, the problem of minimizing the sum of OAs with pilot constraints is formulated as

$$\mathcal{P}_1 : \min_{\mathbf{x}} \sum_{i=1}^{K} \sum_{j=1}^{K} A_{ij} x_i x_j \tag{12}$$

$$s.t. \quad \sum_{i=1}^{K} x_i = U \tag{12a}$$

$$\sum_{j \in \Omega_i} \beta_i g_s x_j \leq P_C \tag{12b}$$

$$\mathbf{x} \in \{0, 1\}^K, \tag{12c}$$

where U represents the number of scheduled users, β_i denotes the weighted factor used to adjust the number of beams allocated to each user, g_s refers to the average number of beams assigned to each user in the system after scheduling, P_C represents the maximum pilot constraint, and $\mathbf{x} = [x_1, x_2, \ldots, x_K] \in \{0, 1\}^K$ with $x_i \in \{0, 1\}$. For a more efficient system, we aim to allocate a total number of beams close to M when the number of scheduled users is high. Conversely, when the number of scheduled users is low, increasing the number of beams allocated to each user beyond a certain point will not improve system performance. Therefore, an upper bound value ξ is set for the number of beams assigned to each scheduled user. Based on these considerations, the value of g_s is set to $\min(\xi, M/U)$. The specific design details of β_i can be found in Sect. 4.

It is evident that the angle matrix $\mathbf{A}$ serves as the coefficient matrix in the objective function of $\mathcal{P}_1$. In convex quadratic programming, the Hessian matrix of the objective function is positive definite. In the case of $\mathcal{P}_1$, the Hessian matrix of the objective function is represented by $\mathbf{L} = 2\mathbf{A}$. If matrix $\mathbf{A}$ is positive definite, then according to the

properties of eigenvalues and the necessary and sufficient conditions for a positive definite matrix, matrix $\mathbf{L}$ is also positive definite.

However, due to the fact that all diagonal elements of matrix $\mathbf{A}$ are 0, some of the sequential principal minors of matrix $\mathbf{A}$ may be smaller than 0. Therefore, matrix $\mathbf{A}$ cannot be a positive definite matrix. To address this, we add a scalar matrix to matrix $\mathbf{A}$, transforming it into a positive definite matrix $\mathbf{A}_P$, which can be expressed as

$$\mathbf{A}_P = \mathbf{A} + \alpha \cdot \mathbf{I}_K. \tag{13}$$

If α surpasses the absolute value of the minimum eigenvalue of matrix $\mathbf{A}$, $\mathbf{A}_P$ is deemed positive definite [36]. For simplicity, we set α as the smallest positive integer that ensures the positive definiteness of matrix $\mathbf{A}_P$. By replacing the coefficient matrix in the objective function of $\mathcal{P}_1$ with matrix $\mathbf{A}_P$, the optimization problem can be transformed into the following matrix form

$$\mathcal{P}_2 : \min_{\mathbf{x}} \mathbf{x}^T \mathbf{A}_P \mathbf{x} \tag{14}$$

$$s.t.\ \ \mathbf{e}^T \mathbf{x} = U \tag{14a}$$

$$\mathbf{n}_B \odot (\hat{\mathbf{A}}_f \mathbf{x}) \leq P_C \cdot \mathbf{e} \tag{14b}$$

$$\mathbf{x} \in \{0, 1\}^K, \tag{14c}$$

where $\mathbf{e} = [1, 1, \ldots, 1]^T \in \mathbb{N}^{K \times 1}$, $\hat{\mathbf{A}}_f$ is a matrix formed by setting the diagonal elements of matrix $\hat{\mathbf{A}}$ to 1, $\mathbf{n}_B = [\beta_1 g_s, \beta_2 g_s, \ldots, \beta_K g_s]^T$, and $\beta_k g_s$ is the number of beams serving users in the neighbor domain of user k. It is worth noting that $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent, as they share the same optimal solution and their optimal values differ by a constant $\alpha \cdot U$.

1 *Remark*

Matrix $\mathbf{A}$ *(or matrix* $\mathbf{A}_P$*) is derived from two directional features of the users, namely the azimuth angle* θ *and angular spread* Δ*. As a result, the proposed user scheduling algorithm only requires these two directional features to perform the scheduling task.*

4 Linearization of 0–1 quadratic programming

In this section, we propose methods to solve β_i and the scheduling problem $\mathcal{P}_2$.

4.1 Beam allocation based on overlap density

Before scheduling, only azimuth angle θ, AS Δ and NAS ω can be determined. It is crucial to note that the pre-beamforming matrix $\mathbf{B}$ of N-JSDM is solved iteratively, and $span(\mathbf{B}) = span^{\perp}(\bar{\mathbf{C}}_k) \bigcap span(\mathbf{C}_k) \bigcap span^{\perp}(\mathbf{B}_{\Psi_k})$, where $\mathbf{B}_{\Psi_k} = [\mathbf{B}_{k_l}, \mathbf{B}_{k_l+1}, \ldots, \mathbf{B}_{k-1}]$. This implies that the azimuth angle of users must be known during the process of solving $\mathbf{B}$, making it challenging to obtain matrix $\mathbf{B}$ during the user scheduling process. Therefore, we propose a beam allocation method based on the overlap density of neighbor

domains. The method is aimed at determining the number of columns of the pre-beamforming matrix $\mathbf{B}_k$.

Note that when the local distribution of users is dense, it is advisable to use a small number of beams to serve these users and use more beams to serve other users. This approach is based on the fact that the number of beams in a particular angle area is limited, and it can not only reduce the pilot overhead but also enable the system to serve more users.

The overlap density of the neighbor domain Ω_k is used to describe the average degree of overlap between any two users in set Ω_k and can be calculated as

$$\rho_k = \frac{\frac{1}{2}\sum_{i,j\in\Omega_k} A_{ij}}{C^2_{|\Omega_k|}\cdot 2\Delta}, \tag{15}$$

where the numerator represents the sum of OAs between users in Ω_k, the coefficient $\frac{1}{2}$ is due to the real symmetry of the angle matrix $\mathbf{A}$, and $C^2_{|\Omega_k|} = \frac{|\Omega_k|!}{2!(|\Omega_k|-2)!}$ in the denominator is the combination number formula. Considering that the OA range between users in set Ω_k is $(0, 2\Delta]$. The denominator of Eq. (15) represents the upper bound of the sum of OAs between users in Ω_k, which is equal to the superposition of the maximum OAs of any two users in Ω_k. It should be noted that the value range of ρ_k is $(0, 1]$.

Next, ρ_k is used to determine the average number of beams allocated to each user within the set Ω_k. As explained in Sect. 3, the average number of beams serving each user in Ω_k is $\beta_k g_s$. Assuming that the value range of $\beta_k g_s$ is $[g_s - \tau, g_s + \tau]$, then the expression of $\beta_k g_s$ is as follows

$$\beta_k g_s = \begin{cases} g_s + \tau(1-2\rho_k) & ,|\Omega_k| \neq 1; \\ g_s & ,|\Omega_k| = 1. \end{cases} \tag{16}$$

As seen in (16), when some users in the system are densely distributed, i.e., the overlap density of their neighbor domains is high, the number of beams serving these users will decrease, and vice versa. The detail of how to obtain $\mathbf{n}_B$ is in Algorithm 1.

1 *Remark*

The value of τ should not be too large, because when the overlap density of the neighbor domain Ω_k is small, the average number of beams serving these users will be close to $g_s + \tau$. This implies that the total number of beams serving users in this neighbor domain will increase by $\tau|\Omega_k|$. Additionally, it is essential to emphasize that while solving problem $\mathcal{P}_2$ (which will later be transformed into problem $\mathcal{P}_5$), we do not have knowledge of the exact number of beams serving each user, but only the average value in the neighbor domain.

Algorithm 1 Beam Allocation Based on Overlap Density

Input: θ_k $(k = 1, 2, \cdots, K)$, Δ, ω, τ, g_s
Output: $\mathbf{n}_B$
Use (8) to get angle matrix $\mathbf{A}$;
for $k = 1 : K$ **do**
    Obtain the edge neighbor users k_l and k_u of user k;
end for
for $k = 1 : K$ **do**
    Use (15) to get the overlap density ρ_k;
    Use (16) to get the average number of beams serving users in the user k's neighbor domain;
end for
return $\mathbf{n}_B$

4.2 Linearization

Note that the user scheduling problem in $\mathcal{P}_2$ is a 0–1 quadratic programming problem whose computational complexity increases exponentially with the problem size. To solve $\mathcal{P}_2$ with a low-computational complexity, we further transform it into a 0–1 mixed integer linear programming as follows. Consider the following optimization problem

$$\mathcal{P}_3 : \min_{\mathbf{x},\mathbf{z},\mathbf{s}} \mathbf{e}^T \mathbf{s} \tag{17}$$

$$s.t.\ \mathbf{A}_P \mathbf{x} - \mathbf{z} - \mathbf{s} = \mathbf{0} \tag{17a}$$

$$\mathbf{z}^T \mathbf{x} = 0 \tag{17b}$$

$$\mathbf{e}^T \mathbf{x} = U \tag{17c}$$

$$\mathbf{n}_B \odot (\hat{\mathbf{A}}_f \mathbf{x}) \leq P_C \cdot \mathbf{e} \tag{17d}$$

$$\mathbf{z} \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}, \mathbf{x} \in \{0, 1\}^K, \tag{17e}$$

where $\mathbf{z} \in \mathbb{R}^{K \times 1}$ and $\mathbf{s} \in \mathbb{R}^{K \times 1}$.

Theorem 1 *$\mathcal{P}_2$ has an optimal solution $\mathbf{x}^*$ if and only if there are $\mathbf{z}^*$ and $\mathbf{s}^*$ such that $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is an optimal solution to $\mathcal{P}_3$, and $\mathcal{P}_2$ and $\mathcal{P}_3$ have the same optimal solution.*

1 ***Proof***

See Appendix A.

It can be observed that the constraint (17b) in $\mathcal{P}_3$ is quadratic, so $\mathcal{P}_3$ is not a linear programming. To further process $\mathcal{P}_3$, we proceed as follows: From (17b), we can deduce that if $x_i = 1$, then z_i must be 0, but if $x_i \neq 1$, then z_i is not necessarily 0. Moreover, from (17a), we have $\mathbf{z} \leq \mathbf{A}_P \mathbf{x}$, implying an upper bound on $\mathbf{z}$. Thus, we have $\mathbf{z} \leq \mathbf{A}_P \mathbf{x} \leq \|\mathbf{A}_P\|_\infty \cdot \mathbf{e}$, where $\|\mathbf{A}_P\|_\infty = \max_i \sum_{j=1}^{K} |a_{ij}|$ is the infinite norm of the matrix $\mathbf{A}_P$. By letting $M_T = \|\mathbf{A}_P\|_\infty$ and using $\mathbf{z} \leq M_T(\mathbf{e} - \mathbf{x})$ to replace $\mathbf{z}^T \mathbf{x} = 0$, we can transform $\mathcal{P}_3$ into the following form

$$\mathcal{P}_4 : \min_{\mathbf{x},\mathbf{z},\mathbf{s}} \mathbf{e}^T\mathbf{s} \tag{18}$$

$$s.t.\ \mathbf{A}_P\mathbf{x} - \mathbf{z} - \mathbf{s} = \mathbf{0} \tag{18a}$$

$$\mathbf{z} \leq M_T(\mathbf{e} - \mathbf{x}) \tag{18b}$$

$$\mathbf{e}^T\mathbf{x} = U \tag{18c}$$

$$\mathbf{n}_B \odot (\hat{\mathbf{A}}_f\mathbf{x}) \leq P_C \cdot \mathbf{e} \tag{18d}$$

$$\mathbf{z} \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}, \mathbf{x} \in \{0,1\}^K. \tag{18e}$$

□

Theorem 2 $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is a feasible solution of* $\mathcal{P}_3$ *if and only if* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is a feasible solution of* $\mathcal{P}_4$; $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is an optimal solution of* $\mathcal{P}_3$ *if and only if* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is an optimal solution of* $\mathcal{P}_4$.

1 ***Proof***

When $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is a feasible solution of* $\mathcal{P}_3$*, obviously* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is a feasible solution of* $\mathcal{P}_4$*. Assuming that* $\mathcal{P}_4$ *has a feasible solution* $(\mathbf{x},\mathbf{z},\mathbf{s})$*, because of* $\mathbf{0} \leq \mathbf{z} \leq M_T(\mathbf{e} - \mathbf{x})$*, when* $x_i = 1$*, there must be* $z_i = 0$*, while* $x_i \neq 1$ *implies that* $\mathbf{z} \leq M_T$*. Therefore, we can obtain* $\mathbf{z}^T\mathbf{x} = 0$*, indicating that* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is also a feasible solution of* $\mathcal{P}_3$*. Similarly, it can be proven that* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is an optimal solution of* $\mathcal{P}_3$ *if and only if* $(\mathbf{x},\mathbf{z},\mathbf{s})$ *is an optimal solution of* $\mathcal{P}_4$*.*

4.3 The algorithm to obtain scheduled users and beams

It is worth noting that the solution space dimension of $\mathcal{P}_4$ is $3K$. This implies that if the number of original users in the system is large, the solution space dimension of $\mathcal{P}_4$ will also be large. As the computational complexity grows with the size of the problem, $\mathcal{P}_4$ still has a high complexity when the user scale is large. Thus, we simplify $\mathcal{P}_4$ as follows: Since $\mathbf{A}_P\mathbf{x} - \mathbf{z} - \mathbf{s} = \mathbf{0} \Leftrightarrow \mathbf{A}_P\mathbf{x} - \mathbf{s} = \mathbf{z}$ and $\mathbf{0} \leq \mathbf{z} \leq M_T(\mathbf{e} - \mathbf{x})$, the constraints (18a) and (18b) can be transformed into $\mathbf{0} \leq \mathbf{A}_P\mathbf{x} - \mathbf{s} \leq M_T(\mathbf{e} - \mathbf{x})$. Hence, $\mathcal{P}_4$ can be transformed into

$$\mathcal{P}_5 : \min_{\mathbf{x},\mathbf{z},\mathbf{s}} \mathbf{e}^T\mathbf{s} \tag{19}$$

$$s.t.\ \mathbf{A}_P\mathbf{x} - \mathbf{s} \geq \mathbf{0} \tag{19a}$$

$$\mathbf{A}_P\mathbf{x} - \mathbf{s} \leq M_T(\mathbf{e} - \mathbf{x}) \tag{19b}$$

$$\mathbf{e}^T\mathbf{x} = U \tag{19c}$$

$$\mathbf{n}_B \odot (\hat{\mathbf{A}}_f \mathbf{x}) \leq P_C \cdot \mathbf{e} \tag{19d}$$

$$\mathbf{s} \geq \mathbf{0}, \mathbf{x} \in \{0,1\}^K. \tag{19e}$$

$\mathcal{P}_5$ is a mixed integer linear programming that can be solved using the branch and bound algorithm. Here, we implemented the branch and bound algorithm using the MOSEK optimization solver [37] in the CVX toolbox.

1 *Remark*

In practical scenarios, local users may be densely distributed, and/or the pilot requirements P_C may be too strict, leading to situations where problem $\mathcal{P}_5$ has no solution. In such cases, we choose to gradually reduce the number of scheduled users U until they can be effectively served. To achieve this, we reduce one user at a time, update g_s, and then recalculate the solution of $\mathcal{P}_5$ based on the updated conditions.

Once $\mathcal{P}_5$ has been solved, we can determine the average number of beams serving each user in each neighbor domain and the scheduled users. However, the exact number of beams serving each user remains unknown. To address this issue, we utilize a linear system of equations to calculate g_k. Firstly, we sort the users in ascending order based on their azimuth angle and obtain the angle matrix $\mathbf{A}_S$ of the scheduled users. We set its diagonal elements to 1 and convert it into an unweighted matrix $\hat{\mathbf{A}}_S$. Then, we sum the rows of the matrix and convert it into a diagonal matrix $\mathbf{D}_S$. The form of matrix $\mathbf{D}_S$ is as follows

$$\mathbf{D}_S = diag(|\Omega_1|, |\Omega_2|, \ldots, |\Omega_U|). \tag{20}$$

We can also get the $\beta_k g_s (k = 1, 2, \ldots, K)$ corresponding to the remaining users and sort them in ascending order, i.e., $\check{\mathbf{n}}_B = (\check{\beta}_1 g_s, \check{\beta}_2 g_s, \ldots, \check{\beta}_U g_s)$. Considering that some users are neighbors with each other but non-neighbors with other users, we take the average of $\check{\beta}_u$ for these neighbor users. The system of equations for solving $g_u (u = 1, 2, \ldots, U)$ is as follows

$$\hat{\mathbf{A}}_S \mathbf{g} = \mathbf{D}_S \check{\mathbf{n}}_B. \tag{21}$$

The solution of $\mathbf{g}$ is $\mathbf{g} = [g_1, g_2, \ldots, g_U]^T = (\hat{\mathbf{A}}_S)^\dagger \mathbf{D}_S \check{\mathbf{n}}_B$. As the solution for g_u may contain decimal values, we perform round down operation on it, i.e., $g_u = \lfloor g_u \rfloor$, and set the solution of g_u to 1 if it is less than 1. Please refer to Algorithm 2 for the details of solving $\mathcal{P}_5$ and determining the number of beams.

Algorithm 2 Acquisition of Scheduling users and the number of beams

Input: θ_k $(k = 1, 2, \cdots, K)$, U, Δ, ω, P_C
Output: $\mathbf{x}$, $\mathbf{g}$
Initialize $\mathbf{x} = \emptyset$;
Obtain angle matrix $\mathbf{A}$;
Obtain unweighted matrix $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}}_f$;
Obtain positive definite matrix $\mathbf{A}_P$;
Obtain infinite norm M_T;
while $\mathbf{x} = \emptyset$ **do**
    Set $g_s = M/U$;
    Use Algorithm 1 to obtain $\mathbf{n}_B$;
    Solve optimization problem $\mathcal{P}_5$;
    Set $U = U - 1$;
end while
Obtain unweighted matrix $\hat{\mathbf{A}}_S$;
Obtain diagonal matrix $\mathbf{D}_S$;
Use (21) to get $\mathbf{g}$;
return $\mathbf{x}$, $\mathbf{g}$

There are a total of three benchmark algorithms considered in this paper. It should be noted that user scheduling and pre-beamforming in the active channel sparsification [22] method are coupled, requiring the solution of a mixed integer linear programming for joint beam and user selection. However, without specifying the optimization toolkit used, it is not possible to determine its computational complexity. Therefore, we conduct a brief analysis and comparison of the computational complexity of proposed algorithm and the other two benchmark algorithms. The user scheduling in conventional JSDM schemes consists of two stages: user grouping and intra-group user scheduling. In the user grouping stage, the computational complexity of the K-means user grouping method with chordal distance in [25] is $\mathcal{O}(N_{it}KG(2M^3 + M^2))$, where N_{it} is the default number of iterations. The computational complexity of the agglomerative user clustering method with chordal distance in [28] is $\mathcal{O}(\frac{K(K-1)}{2}(2M^3 + M^2))$. Since intra-group user scheduling is often coupled with beamforming design, it would be unfair to compare its computational complexity with our proposed algorithm. The computational complexity of the greedy intra-group user scheduling algorithm in [26] is $\mathcal{O}(UK)$ after modifying the termination condition to scheduling U users and the complexity of beamforming design being ignored. It can be observed that the user grouping stage is the main contributor to the complexity.

In contrast, the computational complexity of our proposed algorithm (Algorithm 2) depends on the complexity of two sub-processes: Algorithm 1 and optimization problem $\mathcal{P}_5$. The computational complexity of Algorithm 1 is $\mathcal{O}(K(K-1))$, where $K-1$ is the number of times to determine the edge neighbors of each user. Optimization problem $\mathcal{P}_5$ is solved using the branch and bound method, with a computational complexity of $\mathcal{O}(2^{2K})$, where $2K$ represents the problem scale. Therefore, the overall computational complexity of our proposed algorithm is $\mathcal{O}(U(K(K-1) + 2^{2K}))$.

4.4 Discussion on proposed user scheduling algorithm

This scheme has three advantages. First, it proposes a beam allocation method that considers the overlap density in the neighboring domain, which guarantees that all scheduled users can be served. This is due to the problem that the pre-beamforming design

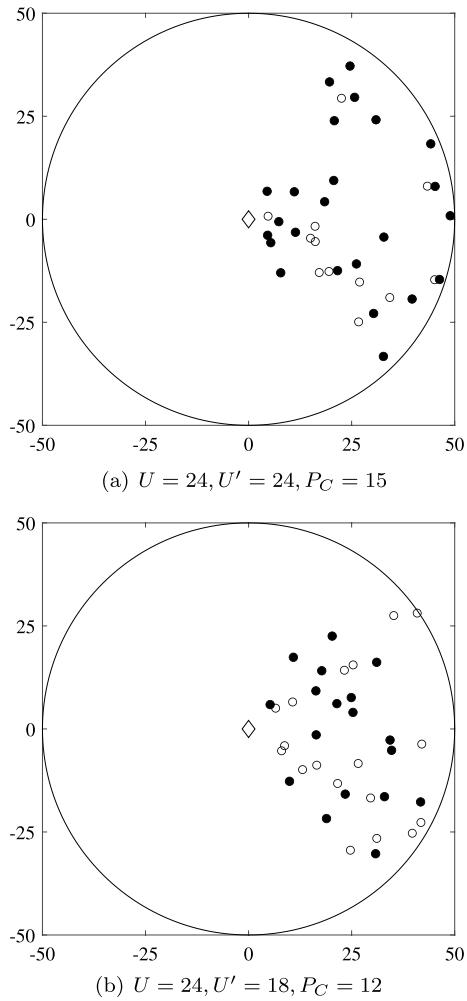


Fig. 1 User scheduling scenarios

method with constrained DTL in N-JSDM cannot be implemented because of the dense local user distribution. Second, the scheme takes into account the influence of the pilot. Furthermore, the scheme is adaptive. If there is no solution to the optimization problem, the number of scheduled users will be gradually reduced until they can be served. Gradually reducing the number of scheduled users in practical scenarios until they can be effectively served brings the following benefits: reduced system load, improved user experience, and decreased interference levels, among others [25].

Figures 1 and 2 illustrate two examples of user scheduling results in a macro-cell scenario. Figures 1a and 2a are drawn from the same initial distribution of users, as are Figs. 1b and 2b. In Fig. 1, the hollow diamond at the center represents the massive MIMO base station, and the large circle indicates the coverage area with a radius of 50 km. Other markers represent users, where hollow circles denote unscheduled users, and solid circles represent scheduled users. In certain scenarios, the actual number of scheduled users, denoted as U', may fall short of our expectations due to unfavorable initial user distributions and stringent pilot conditions (for example, Fig. 1b).

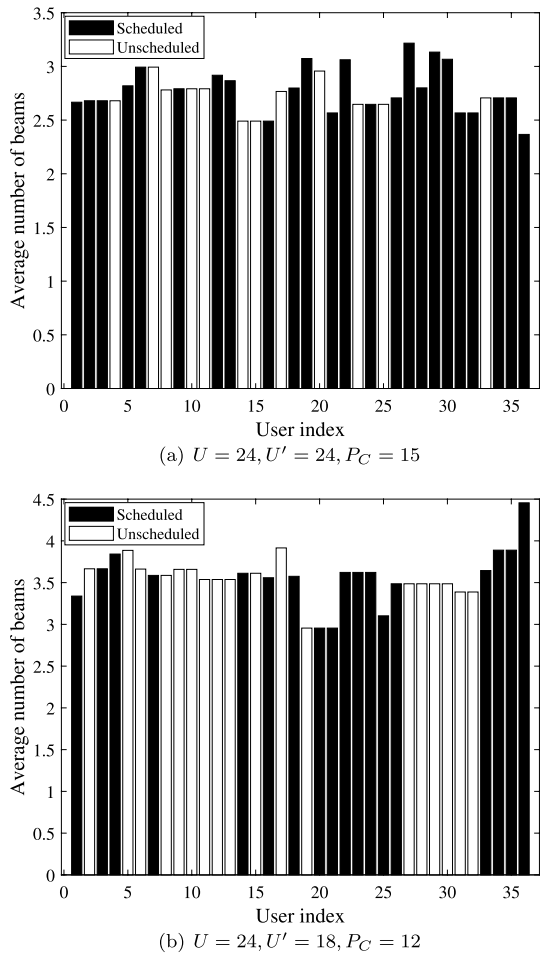


Fig. 2 Average number of beams under different pilot constraints

In Fig. 2, we employ bar graphs to illustrate the scheduling status of users. The vertical axis represents the average number of beams serving each user in their respective neighborhoods, while the horizontal axis represents the user indices. Unfilled bars indicate unscheduled users, while filled bars indicate scheduled users. It can be observed that when the system imposed a limited length of pilots, the desired number of scheduled users cannot be achieved, resulting in $U' < U$. In such cases, the relationship between the average number of beams serving users in their neighborhoods and their scheduling status is not evident. The high average number of beams serving each user in the neighborhood can be attributed to two factors: low overlap density in user neighborhoods and users having fewer neighbors. According to the expression of the pilot (22), we know that the pilot is not only related to the number of neighbors but also to the total number of beams serving users in their neighborhoods. Hence, even if the average number of beams per user is relatively small, a subset of users will still be scheduled to ensure the scheduling of U' users within the limited pilot length. User 16 and user 36 in Fig. 2a and user 20 and user 21 in Fig. 2b serve as examples of this scenario.

Figure 3 displays the ESEs under different P_Cs. As P_C increases, the ESE initially increases and then levels off. This indicates that a small number of P_Cs often leads to a

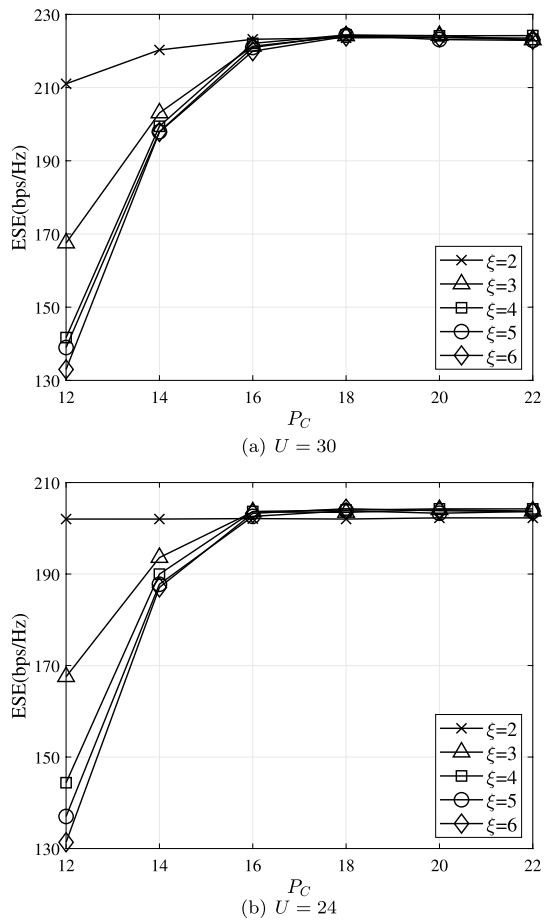


Fig. 3 Comparison of the ESEs under different P_Cs. $K = 36$, $\Delta = 5°$, NAS $= 10°$, SNR $= 20$ dB

failure in scheduling the expected number of users, resulting in performance degradation. Furthermore, it can be observed that a small value of parameter ξ primarily helps to maintain a better level of ESE when P_C is relatively small. Additionally, as shown in Fig. 3b, when the number of scheduled users is small, the value of P_C that results in a smoother ESE will also decrease proportionally. From Fig. 3b, we can also observe that when the number of scheduled users is small and P_C is large, the ESE for $\xi = 2$ is significantly lower than for other values. This discrepancy occurs because a larger P_C value generally leads to a higher likelihood of achieving the expected number of scheduled users. Considering the condition $M/U > 2$, it implies that with $\xi = 2$, fewer beams are allocated to each user compared to other values. This limitation restricts the column number of $\mathbf{B}$ to a significantly smaller value than the number of antennas M, resulting in a larger discrepancy between the column space of $\mathbf{B}$ and the column space of $\mathbf{C}$ compared to other values. Consequently, the ESE is lower for $\xi = 2$. Therefore, the parameter ξ should be set based on both P_C and the number of scheduled users U to optimize system performance.

5 Pre-beamformer design with dynamic beam configuration

We now present the dynamic pre-beamformer design for scheduled users, which differs from previous pre-beamformer designs in N-JSDM by considering the specific user distribution to dynamically configure the beams. The previous designs include the optimal design and the design method with constrained DTL [21]. While the optimal design achieves good performance with a large DTL, it is not suitable for pilot-constrained scenarios. To reduce the DTL, the constrained DTL design limits the number of columns of the pre-beamforming matrix for each user [21]. Specifically, the number of columns of the pre-beamforming matrix $\mathbf{B}$ is set to $\lfloor gK \rfloor$, where $g = M/K$, and $\lfloor \cdot \rfloor$ is the round down operation. The number of columns in $\mathbf{B}_k$ is $\lfloor gk \rfloor - \lfloor g(k-1) \rfloor$.

However, the constrained DTL design method has a fixed number of columns for $\mathbf{B}_k$, which makes it unsuitable for scenarios with randomly distributed users. Therefore, we propose dynamically configuring the number of columns in $\mathbf{B}_k$. Notably, in scenarios with harsh pilot conditions, the optimal design may not meet the transmission requirement, while our proposed method can satisfy it. In the following, we describe how to implement this method using the obtained θ_u and g_u.

Assume that θ_u and g_u of the scheduled users are given. For the unity of symbols, we still use the subscript k to denote the parameters related to user k in this section. From Eq. (4), we can know that user k only needs to feed back $\mathbf{h}_k^H \mathbf{B}_{\Omega_k}$ to BS. The feedback length d_k equals to the number of elements of $\mathbf{h}_k^H \mathbf{B}_{\Omega_k}$, i.e., the number of columns of $\mathbf{B}_{\Omega_k}$. The minimum DTL is $L = \max_k d_k$. In this work, we consider the case where the pilot is the minimum DTL.

In this paper, the pilot length is limited. Since the number of columns of matrix $\mathbf{B}_k$ is g_k, the index set of user k and its neighbors has a linear relationship with the number of columns of $\mathbf{B}_{\Omega_k}$, i.e., the number of neighbors of the user has a linear relationship with the number of pilots. The pilot length P is given by

$$P = \max_{\Omega_k} \sum_{i \in \Omega_k} g_i. \tag{22}$$

To mitigate the non-neighbors' interference of user k, the pre-beamforming matrix $\mathbf{B}$ needs to be designed satisfying Eq. (3). Considering that if user i is a neighbor of user k, then conversely, user k is also a neighbor of user i. Therefore, we can regard Eq. (3) as the problem of designing matrix $\mathbf{B}_k$ to satisfy

$$\mathbf{h}_i^H \mathbf{B}_k = \mathbf{0}, i \in \bar{\Omega}_k, \tag{23}$$

for each k. According to Karhunen–Loeve representation, we can express the channel vector of user k as $\mathbf{h}_k = \mathbf{C}_k^{1/2} \mathbf{z}_k$, where $\mathbf{C}_k^{1/2}$ is a Hermitian matrix. Substituting this into Eq. (23), we obtain the equivalent form

$$\mathbf{z}_i^H \mathbf{C}_i^{1/2} \mathbf{B}_k = \mathbf{0}, i \in \bar{\Omega}_k. \tag{24}$$

During the pre-beamforming stage, only the CCMs $\mathbf{C}_k$ are available at the BS. Without the knowledge of $\mathbf{z}_i$, Eq. (24) can be reformulated as

$$\mathbf{C}_i^{1/2} \mathbf{B}_k = \mathbf{0}, i \in \bar{\Omega}_k. \tag{25}$$

This implies $span(\mathbf{B}_k) \subseteq span^{\perp}(\mathbf{C}_i^{1/2})$ for each $i \in \bar{\Omega}_k$. Based on *Lemma* 1 in [21], we have $span(\mathbf{B}_k) \subseteq span^{\perp}(\bar{\mathbf{C}}_k)$, where $\bar{\mathbf{C}}_k = \sum_{i\in\bar{\Omega}_k} \mathbf{C}_i$.

To fully utilize the signal space and achieve large spectral efficiency, we design $span(\mathbf{B})$ to be close to $span(\mathbf{C})$. This is because the spectral efficiency of the system will be maximized when design $\mathbf{B}$ satisfying $\mathcal{S}_{\mathbf{C}} \cap span(\mathbf{H}) \subseteq span(\mathbf{B})$ where $\mathcal{S}_{\mathbf{C}} = \bigcup span^{\perp}(\bar{\mathbf{C}}_k)$, and $span(\mathbf{H}) \subseteq \bigcup span(\mathbf{C}_k) \subseteq \mathcal{S}_{\mathbf{C}}$ [21]. The difference between two spaces is represented by the chordal distance [25], and the chordal distance of spaces $span(\mathbf{C})$ and $span(\mathbf{B})$ is

$$D_C(span(\mathbf{B}), span(\mathbf{C})) = \| \mathbf{U_B}\mathbf{U_B}^H - \mathbf{U_C}\mathbf{U_C}^H \|_F^2, \tag{26}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{U_C}$ and $\mathbf{U_B}$ are the orthonormal basis of spaces $span(\mathbf{C})$ and $span(\mathbf{B})$, respectively. In order to design $span(\mathbf{B})$ approaching to $span(\mathbf{C})$, the chordal distance between $span(\mathbf{C})$ and $span(\mathbf{B})$ should be minimized, and the problem of designing $\mathbf{B}$ is formalized as

$$\mathcal{P}_6 : \min_{\mathbf{B}} D_C(span(\mathbf{B}), span(\mathbf{C})) \tag{27}$$

$$s.t.\ span(\mathbf{B}_k) \subseteq span^{\perp}(\bar{\mathbf{C}}_k), k = 1, 2, \ldots, K \tag{27a}$$

$$col(\mathbf{B}_{\Omega_k}) \leq P_C \tag{27b}$$

$$\mathbf{B}^H\mathbf{B} = \mathbf{I}, \tag{27c}$$

where constraint (27a) ensures that the effective channel matrix is a band matrix and $\bar{\mathbf{C}}_k = \sum_{i\in\bar{\Omega}_k} \mathbf{C}_i$, constraint (27b) ensures that the design of matrix $\mathbf{B}$ meets the pilot requirements. $\mathcal{P}_6$ is solved iteratively using a greedy algorithm. First, the space $span(\mathbf{C})$ is divided into K subspaces, i.e., $\mathcal{S}_k = span(\mathbf{C}_k), k = 1, 2, \ldots, K$. Then iteratively solves the pre-beamforming matrix $\mathbf{B}_k$ such that the chordal distance between $\mathcal{S}_k$ and $\bigcup_{j=1}^{k} \mathbf{B}_j$ is minimized. When the iteration is complete, $D_C(span(\mathbf{B}), span(\mathbf{C}))$ will be small. In the k -th iteration, the problem of designing $\mathbf{B}_k$ is as follows

$$\mathcal{P}_7 : \min_{\mathbf{B}_k} D_C(span(\mathbf{G}_k), \mathcal{S}_k) \tag{28}$$

$$s.t.\ span(\mathbf{B}_k) \subseteq span^{\perp}(\bar{\mathbf{C}}_k) \tag{28a}$$

$$col(\mathbf{B}_k) = g_k \tag{28b}$$

$$\mathbf{G}_k^H\mathbf{G}_k = \mathbf{I}, \tag{28c}$$

where $\mathbf{G}_k = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_k]$. Setting the number of columns of the matrix $\mathbf{B}_k$ to the obtained value g_k during the user scheduling stage ensures that the actual pilots of the system are less than or equal to P_C. This is because of the constraint (12b) of the user scheduling problem $\mathcal{P}_1$.

Let $\mathbf{U}_{\mathcal{S}_k}$ be the orthogonal basis of $\mathcal{S}_k$. Since $\mathbf{G}_k$ is the orthogonal basis of $span(\mathbf{G}_k)$, we have

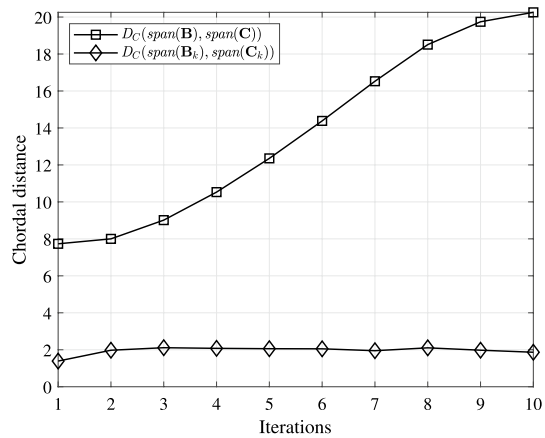


Fig. 4 The chordal distance of different iterations. $K = 20, U = 10, \Delta = 5°$, NAS $= 10°$, SNR $= 20$ dB, $P_C = 20, \xi = 4$

$$\begin{aligned} D_C(span(\mathbf{G}_k), \mathcal{S}_k) =& \| \mathbf{G}_k\mathbf{G}_k^H - \mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H \|_F^2 \\ =& \| \mathbf{G}_{k-1}\mathbf{G}_{k-1}^H + \mathbf{B}_k\mathbf{B}_k^H - \mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H \|_F^2 . \end{aligned} \tag{29}$$

Taking into account the non-negativity property of the Frobenius norm, we now just focus on $\| \mathbf{G}_{k-1}\mathbf{G}_{k-1}^H + \mathbf{B}_k\mathbf{B}_k^H - \mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H \|_F$. Denoting $\mathbf{T} = \mathbf{G}_{k-1}\mathbf{G}_{k-1}^H - \mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H$, we have

$$\begin{aligned} \| \mathbf{B}_k\mathbf{B}_k^H + \mathbf{T} \|_F &= Tr\big((\mathbf{B}_k\mathbf{B}_k^H + \mathbf{T})(\mathbf{B}_k\mathbf{B}_k^H + \mathbf{T})^H\big) \\ &= Tr\big(\mathbf{B}_k\mathbf{B}_k^H\big) + 2Tr\big(\mathbf{B}_k\mathbf{B}_k^H\mathbf{T}\big) + Tr\big(\mathbf{T}\mathbf{T}^H\big). \end{aligned} \tag{30}$$

Based on the property of trace, we can derive $\mathbf{B} = \bar{\mathbf{B}}_{\Psi_k}\mathbf{U}_\varepsilon\mathbf{N}$, where $\bar{\mathbf{B}}_{\Psi_k}$ is the orthogonal basis of the space $span^{\perp}(\mathbf{B}_{\Psi_k})$, and $\mathbf{U}_\varepsilon$ is the matrix composed of the eigenvectors corresponding to the eigenvalues of the matrix $\bar{\mathbf{B}}_{\Psi_k}\bar{\mathbf{R}}_k$ less than ε. For detailed derivations, please refer to [21]. Unlike the design method with constrained DTL, $\mathbf{N}$ is an unitary matrix composed by the eigenvectors of $\mathbf{U}_\varepsilon^H\bar{\mathbf{B}}_{\Psi_k}^H(\mathbf{U}_{\mathcal{S}_k}\mathbf{U}_{\mathcal{S}_k}^H)\bar{\mathbf{B}}_{\Psi_k}\mathbf{U}_\varepsilon$ corresponding to the g_k largest eigenvalues. Once we get $\mathbf{N}$, we can use $\mathbf{B}_k = \bar{\mathbf{B}}_{\Psi_k}\mathbf{U}_\varepsilon\mathbf{N}$ to get $\mathbf{B}_k$.

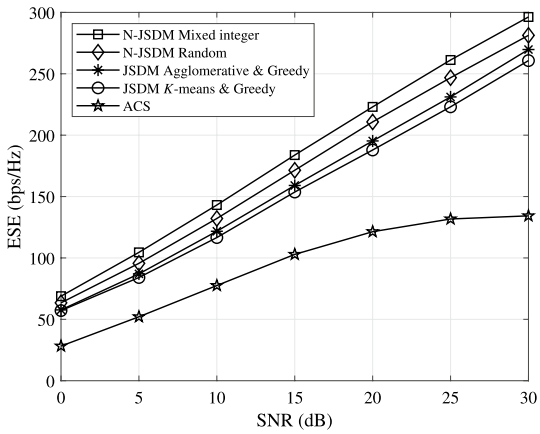
Figure 4 illustrates the chordal distance of different iterations. It should be note that, given the dimension of these two spaces (e.g.,N_1 and N_2), a chordal distance of 0 indicates that the spaces are the same. When the chordal distance reaches its maximum value of $N_1 + N_2$, the spaces are orthogonal to each other. From Fig. 4, we can observe that in each iteration, the chordal distance between $\mathbf{B}_k$ and $\mathbf{C}_k$ remains small but nonzero. This is because $\mathbf{B}_k$ is designed not only to approximate $\mathbf{C}_k$, but also to lie in the null space of the CCM $\bar{\mathbf{C}}_k$. The chordal distance between $span(\mathbf{B})$ and $span(\mathbf{C})$ gradually increases with the number of iterations. This can be attributed to the increasing dimension of $span(\mathbf{B})$ over the course of iterations. Therefore, even though $span(\mathbf{B})$ is designed to approach $span(\mathbf{C})$ incrementally, their chordal distance still increases.

6 Simulation results

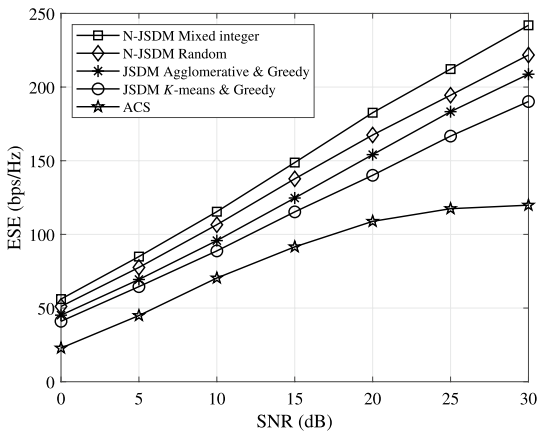
In this section, we provide the simulation results of the proposed algorithm. A ULA with $M = 64$ antennas at the BS is considered, and $K = 36$ single-antenna users are served. The azimuth center angle of each user is uniformly distributed in $[-\frac{\pi}{3}, \frac{\pi}{3}]$ and the angular spread Δ is 5°. For JSDM, the users are partitioned into G groups, and the number of G is proportional to the number of users, i.e., $G = \lfloor K/6 \rfloor$. The value τ in Sect. 4 is set to 1. The parameters in the design of the pre-beamforming matrix $\mathbf{B}$ are consistent with those in [21].

The user scheduling in conventional JSDM consists of two parts: user grouping and intra-group user scheduling. The user grouping stage utilizes the K-means algorithm with chordal distance [25] and the agglomerative user algorithm [28] for grouping users (where the effective channel dimension in the g-th group is $\lfloor M/G \rfloor$). The user scheduling stage employs the algorithm from [26] for scheduling. Given the user grouping, the ESE for scheduled user k in group g is $\eta_{g,k} = (1 - \frac{DTL_{g,k}}{T_C}) \log_2(1 + SINR_{g,k})$, where $SINR_{g,k}$ denotes the SINR of user $k (k = 1, 2, \ldots, K_g)$ in group g and then the overall ESE is $R_{con} = \sum_{g=1}^{G} \sum_{k \in \kappa_g} \eta_{g,k}$.

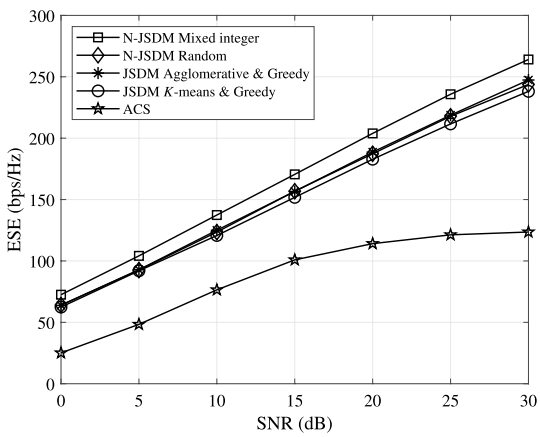
The expected number of active users to be scheduled each time is U. Figure 5 illustrates the ESEs of all algorithms under different SNRs. Our proposed algorithm (denoted by N-JSDM Mixed integer) is compared with two benchmark algorithms of conventional JSDM user scheduling (denoted by JSDM Agglomerative & Greedy and JSDM K-means & Greedy, respectively), the active channel sparsification method (denoted by ACS) as well as N-JSDM with random user scheduling (denoted by N-JSDM Random). The number of scheduled users in Fig. 5a, b is 30, while the number of scheduled users in Fig. 5c is 24. T_C in Fig. 5a, c are 100, and T_C in Fig. 5b is 50. All algorithms exhibit increasing ESE with higher SNR values. It can be seen that our proposed algorithm achieves higher ESE compared to the other algorithms. The performance difference between JSDM Agglomerative & Greedy and JSDM K-means & Greedy stems from their user grouping schemes. The agglomerative user clustering method does not depend on the initial choices of the cluster centers [28]. The ACS consistently exhibits lower ESE compared to other algorithms. This is because the ACS method approximates the downlink CCM of users using the columns of the discrete Fourier transformation matrix compared to other algorithms. This approximation enlarges the energy of both the received signal and interference, and the inter-user interference is directly proportional to the transmission power. When the transmission power is at low level, noise dominates over inter-user interference, and due to the large received power of the signal, the ACS method achieves a large SINR, resulting in a high ESE. However, as the transmission power increases, inter-user interference also increases, leading to no improvement in ESE with increasing SNR. All algorithms except for ACS exhibit similar performance at low SNR. This similarity arises from the fact that in smaller NAS, the impact of DTL on ESE is not significant, and ESE is primarily influenced by spectral efficiency. As the SNR increases, our algorithm achieves higher ESE by minimizing interference and considering spectral overhead. From Fig. 5, it can be observed that the performance gap between N-JSDM and JSDM widens as the number of users increases. This widening gap is attributed to the larger loss of signal space caused by JSDM grouping when the number of users is high, whereas N-JSDM, utilizing the neighbor strategy, can fully leverage the



(a) $U = 30$, $T_C = 100$



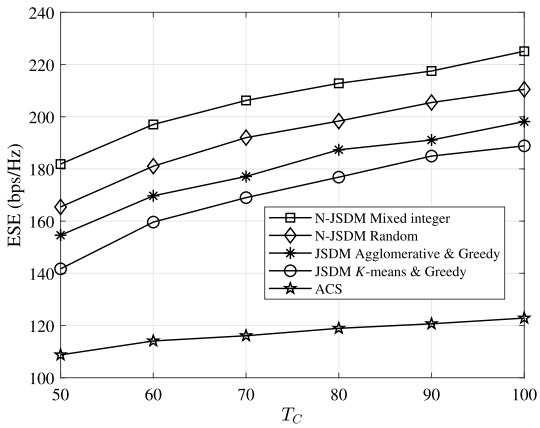
(b) $U = 30$, $T_C = 50$



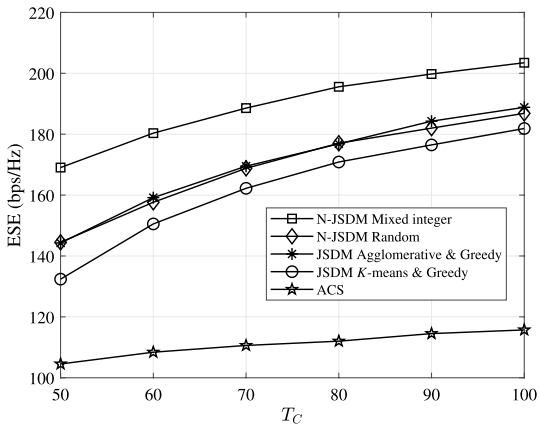
(c) $U = 24$, $T_C = 100$

Fig. 5 Comparison of the ESEs under different SNRs. $K = 36$, $\Delta = 5°$, NAS $= 10°$, $P_C = 20$, $\xi = 4$

signal space, resulting in more significant advantages. Since the ACS method is primarily designed for scenarios where the number of antennas tends to infinity, detailed analysis of ACS performance will not be included in the following simulation.



(a) $U = 30$



(b) $U = 24$

Fig. 6 Comparison of the ESEs under different T_Cs. $K = 36, \Delta = 5°$, NAS $= 10°$, SNR $= 20$ dB, $P_C = 20, \xi = 4$

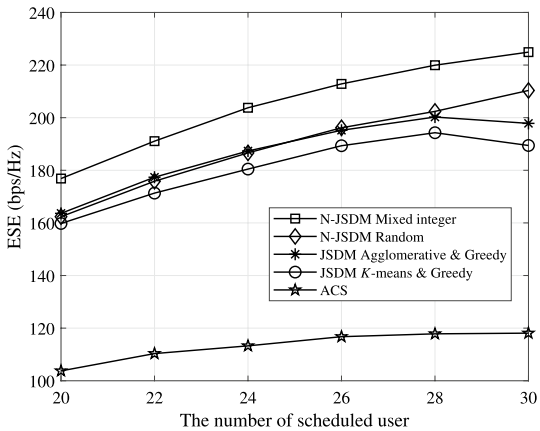


Fig. 7 Comparison of the ESEs under different number of scheduled users. $K = 36, \Delta = 5°$, NAS $= 10°$, SNR $= 20$ dB, $T_C = 100, P_C = 20, \xi = 4$

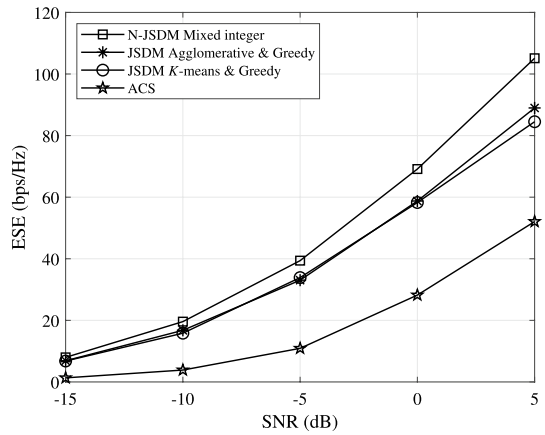


Fig. 8 Comparison of the ESEs under various extremely low SNRs. $K = 36, U = 30, \Delta = 5°$, NAS $= 10°$, $T_C = 100, P_C = 20, \xi = 4$

Figure 6 shows the ESEs of all algorithms under different T_Cs. The weight of spectral overhead in ESE varies with T_C. When T_C is small, the influence of spectral overhead becomes significant since the DTL is positioned in the fractional numerator. As T_C gradually increases, the influence of spectral overhead diminishes, and the significance of spectral efficiency becomes more pronounced. Consequently, the ESE exhibits a gradual upward trend.

Figure 7 depicts the ESEs of all algorithms under different numbers of scheduled users. Several noteworthy observations can be made. Firstly, the ESEs of N-JSDM algorithms increase as the number of scheduled users grows, albeit at a gradually slowing rate. This is because increasing the number of users can enhance spectral efficiency, but it also leads to an increase in interference between users. Secondly, when the number of scheduled users reaches a certain threshold, the performance of conventional JSDM schemes with user scheduling begins to decline. This indicates that as the number of users in the system becomes larger, the performance degradation caused by JSDM grouping becomes more pronounced. Thirdly, due to the approximation used for the CCM, the performance of the ACS method is consistently lower than other algorithms.

In addition, we have conducted additional simulations to evaluate the performance of our proposed algorithm under extreme conditions. These simulations aim to assess the algorithm's robustness and its behavior in challenging scenarios, including scenarios with extremely low SNR and non-uniform user distributions. The performance of the N-JSDM Random is not shown in the following as it is expected that random user scheduling performs inferior to our proposed algorithm. Our focus is on the performance differences between the proposed algorithm and the other benchmark algorithms.

Figure 8 presents the ESEs of all algorithms under different extremely low SNR conditions. It is evident from Fig. 8 that our proposed algorithm consistently achieves higher ESE compared to other algorithms. This superiority stems from our algorithm's scheduling objective of minimizing system interference, which enables better reduction of

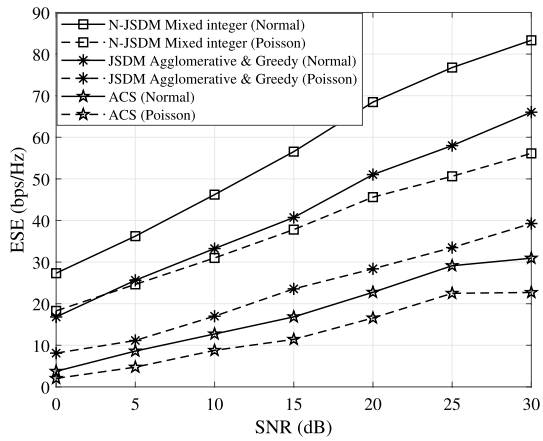


Fig. 9 Comparison of the ESEs under various user distributions. $K = 36, U = 30, \Delta = 5°$, NAS $= 10°$, $T_C = 100, P_C = 20, \xi = 4$

inter-user interference in low SNR scenarios. The performance of JSDM Agglomerative & Greedy and JSDM *K*-means & Greedy is similar, as they both utilize the same scheduling criterion, namely maximizing SINR. The slight performance differences arise from their distinct user grouping methods. On the other hand, the ACS method exhibits the poorest performance due to the approximation employed for the CCM.

Figure 9 illustrates the ESEs of all algorithms under different user distributions, with a standard deviation of 20 for the normal distribution. Comparing it to Fig. 5a, it is evident that the performance of all algorithms experiences a significant decline. This decrease in performance can be attributed to the extreme user distribution, which leads to densely populated local user clusters, making it challenging to achieve the desired number of scheduled users. Furthermore, the interference among the scheduled users is substantial, further contributing to the degradation in performance. To enhance visual clarity, we have omitted the curve for JSDM *K*-means & Greedy, which exhibits marginally lower performance compared to JSDM Agglomerative & Greedy.

7 Conclusion

We proposed a user scheduling method in massive MIMO systems using channel directional characteristics and proposed a dynamic beam allocation method matching the proposed user scheduling. Compared with the complete CSI-based schemes, the two directional features used in this paper, i.e., the azimuth angle and the AS, are generally stable over large time scales. The proposed method scheduled users using mixed integer programming, aiming to improve system performance. Simulations validated the superiority of the proposed method. In our future work, we will extend our method to more channel models, such as Saleh-Valenzuela geometric model and multiple scatterer clusters model.

Appendix

A.1 Proof of Theorem 1

Necessity

Let $\mathbf{x}^*$ *be the optimal solution of* $\mathcal{P}_2$*. Since the elements of matrix* $\mathbf{A}_P$ *are non-negative and* $\forall x_i \in \{0,1\}$*, we have* $\mathbf{A}_P\mathbf{x} \geq \mathbf{0}$*. Thus for* $\mathbf{x}$ *satisfying* $\mathbf{e}^T\mathbf{x} = U$ *and* $\mathbf{n}_B \odot (\hat{\mathbf{A}}_f\mathbf{x}) \leq P_C \cdot \mathbf{e}$*, there must exist* $\mathbf{z} \geq \mathbf{0}, \mathbf{s} \geq \mathbf{0}, \mathbf{z}, \mathbf{s} \in \mathbb{R}^{K\times 1}$ *such that*

$$\mathbf{A}_P\mathbf{x} - \mathbf{z} - \mathbf{s} = \mathbf{0} \tag{17a}$$

$$\mathbf{z}^T\mathbf{x} = 0. \tag{17b}$$

For $\mathbf{x}^*$, $\mathbf{z}^*$ and $\mathbf{s}^*$ satisfying (17a) and (17b), $\mathbf{e}^T\mathbf{s}^*$ is the smallest among all $\mathbf{e}^T\mathbf{s}$.

In the following, we prove that $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution $\mathcal{P}_3$. Use $\mathbf{x}^*$ to replace $\mathbf{x}$ in (17a), and left-multiply $\mathbf{x}^{*T}$ at both ends, we have

$$\mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^* - \mathbf{x}^{*T}\mathbf{z}^* - \mathbf{x}^{*T}\mathbf{s}^* = \mathbf{0}. \tag{31}$$

Due to the constraint $\mathbf{z}^{*T}\mathbf{x}^* = 0$, Eq. (31) is equivalent to

$$\mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^* = \mathbf{x}^{*T}\mathbf{s}^*. \tag{32}$$

Since $\mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^*$ is the smallest among all $\mathbf{x}^T\mathbf{A}_P\mathbf{x}$, $\mathbf{x}^{*T}\mathbf{s}^*$ is also the smallest among all $\mathbf{x}^T\mathbf{s}$. If it can be proved

$$\mathbf{x}^{*T}\mathbf{s}^* = \mathbf{e}^T\mathbf{s}^*, \tag{33}$$

since $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ satisfies the constraints of $\mathcal{P}_3$, $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution of $\mathcal{P}_3$, and $\mathcal{P}_2$ and $\mathcal{P}_3$ have the same optimal value.

Equation (33) is proved in the following. First, it can be shown that there must be $s_i^* = 0$ for $\forall i$ satisfying $x_i^* = 0$. Assuming this does not hold, then there exists some i_s such that when $x_{i_s}^* = 0$, $s_{i_s}^* > 0$, and consequently, $\mathbf{e}^T\mathbf{s}^*$ is the smallest of all $\mathbf{e}^T\mathbf{s}$. Define new $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{s}}$. For $j = 1, 2, \ldots, K$, when $j = i_s$, let $\tilde{z}_j = z_{i_s}^* + s_{i_s}^*, \tilde{s}_j = 0$; when $j \neq i_s$, let $\tilde{z}_j = z_j^*, \tilde{s}_j = s_j^*$. Since $\tilde{\mathbf{z}} + \tilde{\mathbf{s}} = \mathbf{z}^* + \mathbf{s}^*$, then $(\mathbf{x}^*, \tilde{\mathbf{z}}, \tilde{\mathbf{s}})$ also satisfies (17a) and (17b), but $\mathbf{e}^T\tilde{\mathbf{s}} < \mathbf{e}^T\mathbf{s}^*$, which contradicts the method of choosing $\mathbf{s}^*$. Thus, there must be $s_i^* = 0$ for $\forall i$ satisfying $x_i^* = 0$, and Eq. (33) holds. As a result, $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution of $\mathcal{P}_3$, and $\mathcal{P}_2$ and $\mathcal{P}_3$ have the same optimal value.

sufficiency : In the following, we prove that if $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution of $\mathcal{P}_3$, then $\mathbf{x}^*$ is the optimal solution of $\mathcal{P}_2$. We use the contradiction method to complete the proof.

Assuming that $\mathbf{x}^*$ is not the optimal solution of $\mathcal{P}_2$, and $\bar{\mathbf{x}}$ is the optimal solution of $\mathcal{P}_2$, then $\bar{\mathbf{x}}^T\mathbf{A}_P\bar{\mathbf{x}} < \mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^*$. Since $\bar{\mathbf{x}}$ is the optimal solution of $\mathcal{P}_2$, according to the method of finding the optimal solution of $\mathcal{P}_3$ in necessity, $\bar{\mathbf{z}}$ and $\bar{\mathbf{s}}$ satisfying (17a) and (17b) can be obtained, and $\mathbf{e}^T\bar{\mathbf{s}}$ is minimized. From the proof of necessity, it can be known that $(\bar{\mathbf{x}}, \bar{\mathbf{z}}, \bar{\mathbf{s}})$ is the optimal solution of $\mathcal{P}_3$ and satisfies

$$\bar{\mathbf{x}}^T\mathbf{A}_P\bar{\mathbf{x}} = \bar{\mathbf{x}}^T\bar{\mathbf{s}} = \mathbf{e}^T\bar{\mathbf{s}}. \tag{34}$$

However, since $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution of $\mathcal{P}_3$, it follows from the proof of necessity that

$$\mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^* = \mathbf{x}^{*T}\mathbf{s}^* = \mathbf{e}^T\mathbf{s}^*. \tag{35}$$

Since $\bar{\mathbf{x}}^T\mathbf{A}_P\bar{\mathbf{x}} < \mathbf{x}^{*T}\mathbf{A}_P\mathbf{x}^*, \mathbf{e}^T\mathbf{s}^* > \mathbf{e}^T\bar{\mathbf{s}}$ can be obtained, which contradicts that $(\mathbf{x}^*, \mathbf{z}^*, \mathbf{s}^*)$ is the optimal solution of $\mathcal{P}_3$. Hence, *Theorem* 1 is proved.

Abbreviations

MIMO	Multiple-input multiple-output
BS	Base station
CSI	Channel state information
TDD	Time division duplex
FDD	Frequency division duplex
DTL	Downlink training length
JSDM	Joint spatial division multiplexing
CCM	Channel covariance matrix
SINR	Signal-to-interference-plus-noise ratio
SLNR	Signal-to-leakage-plus-noise ratio
N-JSDM	Neighbor-based JSDM
ASR	Angle-spreading-ranges
CQI	Channel quality indicator
ESE	Effective spectral efficiency
ULA	Uniform linear array
OR	One-ring
AS	Angular spread
NAS	Neighbor angular spread
IUI	Inter-user interference
OA	Overlap angle

Acknowledgements
The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

Author contributions
HL conducted research, conceptualized, simulated and modified experiments, as well as wrote the manuscript; CL and YS helped to conceive the idea and revise the manuscript; TG suggested improvements and revised the manuscript; YZ assisted with data handling and manuscript revisions. All authors read and approved the final manuscript.

Funding
This work was supported in part by the Natural Science Foundation of China under Grants 61771257, 62101282, and 62371249.

Availability of data and materials
The datasets simulated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 21 August 2023 Accepted: 14 December 2023
Published online: 02 January 2024

References

1. I.F. Akyildiz, J.M. Jornet, Realizing ultra-massive MIMO (1024 × 1024) communication in the (0.06–10) Terahertz band. Nano Commun. Netw. **8**, 46–54 (2016). https://doi.org/10.1016/j.nancom.2016.02.001
2. R. Hussain, M.S. Sharawi, 5G MIMO antenna designs for base station and user equipment: some recent developments and trends. IEEE Antennas Propag. Mag. **64**(3), 95–107 (2022). https://doi.org/10.1109/MAP.2021.3089983

3. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010). https://doi.org/10.1109/TWC.2010.092810.091092
4. Z. Gao, L. Dai, W. Dai, Z. Wang, Block compressive channel estimation and feedback for FDD massive MIMO. in *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 49–50 (2015). https://doi.org/10.1109/INFCOMW.2015.7179337
5. B. Lee et al., Antenna grouping based feedback compression for FDD-based massive MIMO systems. IEEE Trans. Commun. **63**(9), 3261–3274 (2015). https://doi.org/10.1109/TCOMM.2015.2460743
6. D. Fan, F. Gao, G. Wang, Z. Zhong, A. Nallanathan, Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems. IEEE J. Sel. Areas Commun. **35**(9), 1948–1961 (2017). https://doi.org/10.1109/JSAC.2017.2720938
7. X. Yang et al., Design and implementation of a tdd-based 128-antenna massive MIMO prototype system. China Commun. **14**(12), 162–187 (2017). https://doi.org/10.1109/CC.2017.8246333
8. X. Jiang, F. Kaltenberger, L. Deneire, How accurately should we calibrate a Massive MIMO TDD system? in *2016 IEEE International Conference on Communications Workshops (ICC)*, 706–711 (2016). https://doi.org/10.1109/ICCW.2016.7503870
9. D. Liu, W. Ma, S. Shao, Y. Shen, Y. Tang, Performance analysis of TDD reciprocity calibration for massive MU-MIMO systems with ZF beamforming. IEEE Commun. Lett. **20**(1), 113–116 (2016). https://doi.org/10.1109/LCOMM.2015.2499283
10. Global Market Monitor: China TDD and FDD Spectrum Industry Market Research Report 2023–2029 (2023), https://www.globalmarketmonitor.com/reports/1775253-tdd-and-fdd-spectrum-market-report.html
11. A. Abdallah, M.M. Mansour, Efficient angle-domain processing for FDD-based cell-free massive MIMO systems. IEEE Trans. Commun. **68**(4), 2188–2203 (2020). https://doi.org/10.1109/TCOMM.2020.2969351
12. A. Adhikary, J. Nam, J.-Y. Ahn, G. Caire, Joint spatial division and multiplexing-the large-scale array regime. IEEE Trans. Inf. Theory **59**(10), 6441–6463 (2013). https://doi.org/10.1109/TIT.2013.2269476
13. H.-W. Liang, W.-H. Chung, S.-Y. Kuo, FDD-RT: a simple CSI acquisition technique via channel reciprocity for FDD massive MIMO downlink. IEEE Syst. J. **12**(1), 714–724 (2018). https://doi.org/10.1109/JSYST.2016.2556222
14. Z. Gao, L. Dai, Z. Wang, S. Chen, Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO. IEEE Trans. Signal Process. **63**(23), 6169–6183 (2015). https://doi.org/10.1109/TSP.2015.2463260
15. Y. Ding, B.D. Rao, Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems. IEEE Trans. Wireless Commun. **8**(17), 5437–5451 (2018). https://doi.org/10.1109/TWC.2018.2843786
16. S. Noh, M.D. Zoltowski, Y. Sung, Love, J. David, Pilot beam pattern design for channel estimation in massive MIMO systems. IEEE J. Sel. Topics Signal Process. **8**(5), 787–801 (2014). https://doi.org/10.1109/JSTSP.2014.2327572
17. H. Ren et al., Long-term CSI-based design for RIS-aided multiuser MISO systems exploiting deep reinforcement learning. IEEE Commun. Lett. **26**(3), 567–571 (2022). https://doi.org/10.1109/LCOMM.2021.3140155
18. Z. Peng et al., Analysis and optimization for RIS-aided multi-pair communications relying on statistical CSI. IEEE Trans. Veh. Technol. **70**(4), 3897–3901 (2021). https://doi.org/10.1109/TVT.2021.3062710
19. D. Kim, G. Lee, Y. Sung, Two-stage beamformer design for massive MIMO downlink by trace quotient formulation. IEEE Trans. Commun. **63**(6), 2200–2211 (2015). https://doi.org/10.1109/TCOMM.2015.2429646
20. Y. Jeon et al., New beamforming designs for joint spatial division and multiplexing in large-scale MISO multi-user systems. IEEE Trans. Wirel. Commun. **16**(5), 3029–3041 (2017). https://doi.org/10.1109/TWC.2017.2673845
21. Y. Song et al., Joint spatial division and multiplexing in massive MIMO: a neighbor-based approach. IEEE Trans. Wirel. Commun. **19**(11), 7392–7406 (2020). https://doi.org/10.1109/TWC.2020.3011101
22. M.B. Khalilsarai, S. Haghighatshoar, X. Yi, G. Caire, FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification. IEEE Trans. Wirel. Commun. **18**(1), 121–135 (2019). https://doi.org/10.1109/TWC.2018.2877684
23. W. Tang, Y. Teng, Y. Man, M. Song, Analysis of two-stage precoding schemes for massive multi-user MIMO downlink systems. in *2016 IEEE International Conference on Communication Systems (ICCS)*, 1–6 (2016). https://doi.org/10.1109/ICCS.2016.7833659
24. J. Ma, S. Zhang, H. Li, N. Zhao, V.C.M. Leung, Base station selection for massive MIMO networks with two-stage precoding. IEEE Wirel. Commun. Lett. **6**(5), 598–601 (2017). https://doi.org/10.1109/LWC.2017.2720662
25. J. Nam, A. Adhikary, J.-Y. Ahn, G. Caire, Joint spatial division and multiplexing: opportunistic beamforming, user grouping and simplified downlink scheduling. IEEE J. Sel. Top. Signal Process. **8**(5), 876–890 (2014). https://doi.org/10.1109/JSTSP.2014.2313808
26. Y. Xu, G. Yue, N. Prasad, S. Rangarajan, S. Mao, User grouping and scheduling for large scale MIMO systems with two-stage precoding. in *2014 IEEE International Conference on Communications (ICC)*, 5197–5202(2014). https://doi.org/10.1109/ICC.2014.6884146
27. J. Nam, Y.-J. Ko, J. Ha, User grouping of two-stage MU-MIMO precoding for clustered user geometry. IEEE Commun. Lett. **19**(8), 1458–1461 (2015). https://doi.org/10.1109/LCOMM.2015.2445349
28. X. Sun, X. Gao, G.Y. Li, W. Han, Agglomerative user clustering and downlink group scheduling for FDD massive MIMO systems. in *2017 IEEE International Conference on Communications (ICC)*, 1–6 (2017). https://doi.org/10.1109/ICC.2017.7997313
29. Z. Jiang, A.F. Molisch, G. Caire, Z. Niu, Achievable rates of FDD massive MIMO systems with spatial channel correlation. IEEE Trans. Wirel. Commun. **14**(5), 2868–2882 (2015). https://doi.org/10.1109/TWC.2015.2396058
30. H. Xie, F. Gao, S. Zhang, S. Jin, A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model. IEEE Trans. Veh. Technol. **66**(4), 3170–3184 (2017). https://doi.org/10.1109/TVT.2016.2594706
31. L. Miretti, R.L.G. Cavalcante, S. Stanczak, FDD massive MIMO channel spatial covariance conversion using projection methods. in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3609–3613 (2018). https://doi.org/10.1109/ICASSP.2018.8462048
32. X. Li, S. Jin, H.A. Suraweera, J. Hou, X. Gao, Statistical 3-D beamforming for large-scale MIMO downlink systems over Rician fading channels. IEEE Trans. Commun. **64**(4), 1529–1543 (2016). https://doi.org/10.1109/TCOMM.2016.2530796

33. Y. Huang, L. Yang, M. Bengtsson, B. Ottersten, Exploiting long-term channel correlation in limited feedback SDMA through channel phase codebook. IEEE Trans. Signal Process. **59**(3), 1217–1228 (2011). https://doi.org/10.1109/TSP.2010.2094190
34. Y. Song et al., Domain selective precoding in 3-D massive MIMO systems. IEEE J. Sel. Top. Signal Process. **13**(5), 1103–1118 (2019). https://doi.org/10.1109/JSTSP.2019.2930889
35. N. Omaki, K. Kitao, K. Saito, T. Imai, Y. Okumura, Experimental study on elevation directional channel properties to evaluate performance of 3D-mimo at base station in microcell outdoor to indoor environment. in *2014 IEEE International Workshop on Electromagnetics (iWEM)*. 219–220 (2014). https://doi.org/10.1109/iWEM.2014.6963715
36. S.P. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, UK, 2004)
37. E.D. Andersen, K.D. Andersen, *The MOSEK Interior Point Optimizer for Linear Programming: An Implementation of the Homogeneous Algorithm*. Springer, Boston, 197–232 (2000)

Publisher's Note