Signal Processing 158 (2019) 141–149

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

User scheduling for capacity-Jain's fairness tradeoff in millimeter-wave MIMO systems

Anzhong Hu

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

ARTICLE INFO

Article history:
Received 2 September 2018
Revised 3 December 2018
Accepted 12 January 2019
Available online 15 January 2019

Keywords:
MIMO systems
Beams
Millimeter-wave communication
Scheduling

ABSTRACT

This paper investigates user scheduling in millimeter-wave multiple-input multiple-output systems that can achieve a tradeoff between capacity and Jain's fairness index. By decomposing the original problem into digital beamforming design, duration allocation, and analog beamforming design, each subproblem is solvable. It is proved that the duration allocation problem possesses a property such that the convex optimization algorithm can be employed for the near optimal result. Moreover, the analog beamforming vectors and the user grouping are designed to avoid interference caused by the reuse of the same subspace. Finally, the computational complexity of the proposed approach is analyzed and is compared with that of other approaches. The simulations verify that the proposed approach can achieve the highest fairness and a higher capacity than approaches of close fairness.

1. Introduction

With the development of wireless systems, it is expected that spectral efficiency and capacity will be improved significantly [1,2]. In order to achieve this, the millimeter-wave (mm-wave) multiple-input multiple-output (MIMO) system [3] has been advocated and has attracts many researchers. Owing to the wide bandwidth and the high beamforming gain, the mm-wave MIMO system has great potential in meeting the future demands on wireless systems. Besides, mm-wave MIMO systems are feasible to employ large arrays, which are also massive MIMO systems [4–6].

Although the mm-wave MIMO system is competitive, similar to other wireless systems, it is also confronted with the problem of user scheduling since the number of users is likely to surpass the system restriction. As known to all, user scheduling has been extensively studied. Undoubtedly, the user scheduling approaches should balance between performance and fairness. When performance is emphasized, the greedy algorithm which searches for the candidate of the highest performance each time is usually resorted to [7–11]. In order to reduce the number of candidates, semi-orthogonal user selection that avoids high interference can be employed [12,13]. When fairness is of more importance, the round-robin (RR) scheduling gives each candidate equal opportunity to be scheduled [14,15]. Besides, the Jain's fairness index [16] is also widely adopted. When trading off between performance and fairness, the proportional fair (PF) criterion [17,18] and the max-min criterion [19–21] attracted the most research interests. Briefly, the former tries to maximize the ratio between the current and the average capacity, and the latter tries to maximize the minimum average capacity. Additionally, [22] balances between performance and the Jain's fairness index for the optimal efficiency-Jain tradeoff (EJT) defined therein.

As far as the author knows, user scheduling is rarely investigated for mm-wave MIMO systems. Instead, researcher mainly focus on user scheduling in mm-wave networks. Chiefly, as multiple nodes need to transmit to each other, the scheduling of the concurrent transmission nodes for the whole network is studied in [23–26]. In mm-wave MIMO systems, the limited number of radio frequency (RF) paths in the base station (BS) restricts the number of simultaneously served mobile stations (MSs). Moreover, the channels are sparse since they are constituted by scarce propagation paths. By taking into account this property, authors in [27] propose to schedule MSs whose channel subspaces are non-overlapped or orthogonal. However, neither the hybrid precoding structure nor the beam selection at the MS has been discussed in this article, which are naturally integrated in the scheduling problem. What's more, how to trade off between performance and fairness in mm-wave MIMO systems has not yet been addressed.

In this paper, user scheduling for mm-wave MIMO systems is investigated. The goal is to achieve the optimal EJT, which is taken as a proper criterion for the tradeoff between the capacity and the Jain's fairness index. In regards to the hybrid structure at both the BS and the MS as well as the limited number of RF paths, the user scheduling can be modeled as an integrated optimization problem. By separating the design of the analog and the digital

E-mail address: huaz@hdu.edu.cn

https://doi.org/10.1016/j.sigpro.2019.01.009
0165-1684/

precoders, the latter can be designed for achieving maximum capacity. Then, the analog beams at both sides and the scheduling of the MSs are of finite choices. With the aid of the monotonic trade-off property (MTP) given in [22], it is proven that the investigated problem can be transformed to possess the MTP. Consequently, a convex optimization algorithm is proposed to achieve the optimal EJT. Furthermore, by resorting to the beamspace idea, the subspace of each cluster can be represented by a bin index, with which the beams are selected and the MSs are grouped for simplifying the user scheduling algorithm. The main contributions of this paper are four-fold.

1) The complex scheduling problem is decomposed into several subproblems. First, the digital precoding is designed given the analog precoding and the scheduled MSs. Second, the scheduling duration for each combination of the analog precoder and the user group is solved. Finally, the analog precoding and the user grouping are designed.
2) The possess of the MTP is proved and a convex optimization algorithm is proposed. With minor adjustment on the original problem, it is proved that the scheduling problem possesses the MTP. Then, classical convex optimization methods are merged into an optimization algorithm to solve the problem in low complexity. Above all, the optimal EJT can be achieved.
3) The beams are selected and the users are grouped in a simple way. First, the channel space is divided into subspaces according to the beamspace idea, and the subspace of each cluster is represented by an index related to the angular domain of that cluster. Second, severe interference is avoided by jointly choosing the transmission beam, the receiving beam, and the users in each group. This can further simplify the proposed scheduling algorithm while keep the scheduling result close to the optimal EJT.
4) The computational complexities of the existing approaches and the proposed approach are analyzed. It is shown that the proposed approach is of comparable complexity to most of the existing approaches.

This paper is organized as follows. In Section 2, the system model, the assumptions, and the problem are given. Section 3 presents the decomposition of the problem, the proof of MTP, and the proposed optimization algorithm. In Section 4, the joint optimization of the beams and user groups are depicted. The computational complexities of the methods are analyzed and compared. Section 5 gives the simulation parameters and the numerical results. Finally, conclusions are drawn in Section 6.

Notations: Lower-case (upper-case) boldface symbols denote vectors (matrices); $\mathbf{I}_K$ represents the $K \times K$ identity matrix; $(\cdot)^H$ and $\mathbb{E}\{\cdot\}$ denote the conjugate transpose and the expectation, respectively; $[\cdot]_j$ is the jth element of a vector; $[\cdot]_{:,j}$, $[\cdot]_{j,:}$, and $[\cdot]_{j,k}$ are the jth column, the jth row, and the element in the jth row and kth column of a matrix, respectively; $\mathrm{tr}\{\cdot\}$ is the trace of a matrix; $\mathcal{A}\backslash\mathcal{B}$ is the relative complement of $\mathcal{B}$ in $\mathcal{A}$; $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$; $\mathbf{0}_K$ and $\mathbf{1}_K$ are length-K all-zero and all-one column vectors, respectively; $\mathcal{CN}(\mathbf{0}_K, \mathbf{R})$ represents the zero-mean complex Gaussian distribution with covariance matrix $\mathbf{R}$ for a length-K random column vector; $\succeq$ denotes element-wise inequality $\geq$; $[(\mathbf{A})^+]_{j,l} = \max([\mathbf{A}]_{j,l}, 0)$; $\lfloor\cdot\rfloor$ represents the closest integer that is smaller than the variable; $\mathrm{vec}(\mathcal{A})$ concatenates the elements in $\mathcal{A}$ into a column vector; and i is the imaginary unit.

2. System model

In this section, the wideband mm-wave MIMO system model is presented, and the specific characteristics of the corresponding channel are illustrated. Additionally, the basic assumptions are also described and the problem to be investigated is formed.

As shown in Fig. 1, the wideband mm-wave MIMO system considered consists of one BS and K MSs. The MSs distribute randomly in front of the BS array. The BS is equipped with one uniform rectangular array (URA) of N_B antennas and each MS has one URA of N_M antennas. Additionally, the BS has N_R RF paths, each of which is composed of one analog-to-digital converter (ADC)/digital-to-analog converter (DAC) and one RF chain. Because of the space and the power constraints at the BS, each RF path should be shared by multiple antennas [3,28]. Hence, we have $N_\mathrm{R} < N_\mathrm{B}$. Since the space and the power constraints are more severe at each MS, it is assumed that each MS has only one RF path. Note that the case of more than one RF path at each MS is also practical, to which the scheme and the analysis presented in this paper can also be extended.

2.1. Signal model

Owing to the frequency selective fading in wideband systems, the orthogonal frequency division multiplexing (OFDM) scheme is employed here, as those in [29–34]. Here, we focus on the downlink transmission, and the uplink transmission can be tackled in a similar way. The BS transmits N_S OFDM symbol blocks simultaneously to the N_S MSs, and each symbol block is of length N. As the number of data streams, i.e., N_S, cannot surpass the number of RF chains at either the transmitter or the receivers, we assume $N_\mathrm{S} = N_\mathrm{R}$ and $N_\mathrm{S} < K$, which means only the partial MSs can be simultaneously served by the BS. At the BS, each transmitted symbol block is composed of N symbols, each of which occupies one subcarrier. Thus, there are $N_\mathrm{S}N$ symbols simultaneously transmitted at the BS.

At the tth slot and the nth subcarrier, the N_S transmitted symbols are formed into a vector, $\mathbf{s}_t[n] \in \mathbb{C}^{N_\mathrm{S}\times 1}$, which satisfies $\mathbb{E}\{\mathbf{s}_t[n]\mathbf{s}_t[n]^H\} = \mathbf{I}_{N_\mathrm{S}}$. This symbol vector is first processed with baseband digital precoding. Denote the precoding matrix as $\mathbf{F}_t[n] \in \mathbb{C}^{N_\mathrm{S}\times N_\mathrm{S}}$, the precoded symbol vector is $\mathbf{F}_t[n]\mathbf{s}_t[n] \in \mathbb{C}^{N_\mathrm{S}\times 1}$. Then, each of the N_S precoded symbols goes through one RF path. On each RF path, there are N precoded symbols in total, and are transformed to the time domain with an N-point inverse fast Fourier transform (IFFT). After that, the cyclic prefixes are added on each RF path, i.e., for each symbol block of length N. After passing the RF chains, the symbols are processed with the RF band analog precoding. Here, the analog precoding matrix is the same for all the subcarriers and is denoted as $\mathbf{F}_t^\mathrm{R} \in \mathbb{C}^{N_\mathrm{B}\times N_\mathrm{S}}$. Hence, for the nth subcarrier at the tth slot, the transmitted symbol vector at the BS is given by $\mathbf{x}_t[n] = \mathbf{F}_t^\mathrm{R}\mathbf{F}_t[n]\mathbf{s}_t[n] \in \mathbb{C}^{N_\mathrm{B}\times 1}$. In this paper, the fully-connected structure is assumed, which means that each RF chain is connected with every antenna through a phase shifter. Moreover, the entries of $\mathbf{F}_t^\mathrm{R}$ are restricted to be of constant amplitude with $|[\mathbf{F}_t^\mathrm{R}]_{l,m}| = 1/\sqrt{N_\mathrm{B}}, \forall l, m$. The hybrid precoding power per subcarrier is restricted as $\mathbb{E}\{||\mathbf{x}_t[n]||_\mathrm{F}^2\} = ||\mathbf{F}_t^\mathrm{R}\mathbf{F}_t[n]||_\mathrm{F}^2 \leq P$, where P is the power constraint per subcarrier.

The received signal at the kth MS, the nth subcarrier, and the tth slot can be expressed as $\mathbf{H}_k[n]\mathbf{x}_t[n] + \mathbf{z}_{t,k}[n] \in \mathbb{C}^{N_\mathrm{M}\times 1}$, where $\mathbf{H}_k[n] \in \mathbb{C}^{N_\mathrm{M}\times N_\mathrm{B}}$ is the channel matrix between the BS and the kth MS at subcarrier n, and $\mathbf{z}_{t,k}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_\mathrm{M}})$ is the received noise vector at the nth subcarrier. The kth MS then processes the received symbols with an analog combiner, $\mathbf{w}_{t,k} \in \mathbb{C}^{N_\mathrm{M}\times 1}$, which is also frequency flat, and this yields the symbol, $\mathbf{w}_{t,k}^H\mathbf{H}_k[n]\mathbf{x}_t[n] + \mathbf{w}_{t,k}^H\mathbf{z}_{t,k}[n]$. Note that the phase shifter structure at the MS is in the same form as that at the BS, and entries of the analog precoder are constrained with constant norm as $|[\mathbf{w}_{t,k}]_l| = 1/\sqrt{N_\mathrm{M}}, \forall l$. Then, the cyclic prefixes are removed and the fast Fourier transform (FFT) is employed to change the received symbols into the frequency domain. Finally, the received symbol at the kth MS, the nth

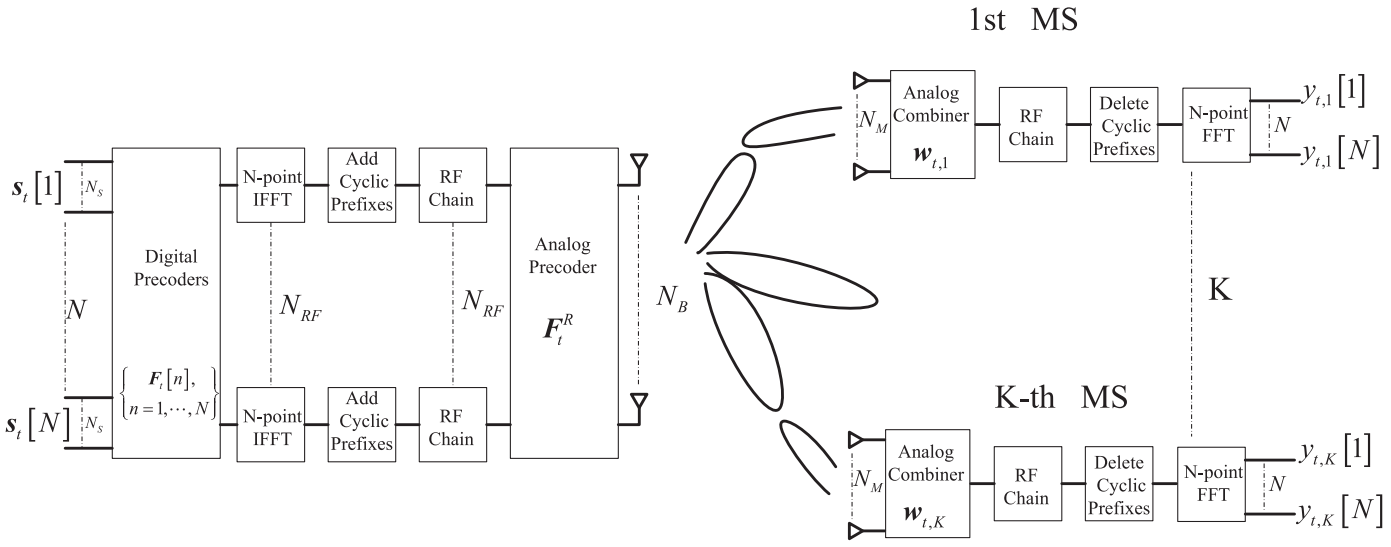

Fig. 1. System model.

subcarrier, and the tth slot can be written as

$$y_{t,k}[n] = \mathbf{w}_{t,k}^H \mathbf{H}_k[n]\mathbf{F}_t^{\mathrm{R}}\mathbf{F}_t[n]\mathbf{s}_t[n] + \mathbf{w}_{t,k}^H \mathbf{z}_{t,k}[n]. \quad (1)$$

Note that all the subcarriers are used for each MS.

2.2. Channel model

As usual, the wideband mm-wave propagation environment is modeled in a geometrical way, with N_{C} clusters of scatterers and N_{SC} scatterers in each cluster. Additionally, the channel is assumed to undergo block fading, and the channel coherence interval is T when measured with slots. Then, the channel matrix between the kth MS and the BS at the nth subcarrier can be expressed as [29–35]

$$\mathbf{H}_k[n] = \sum_{l=1}^{N_{\mathrm{C}}}\sum_{l_c=1}^{N_{\mathrm{SC}}} \alpha_{k,l,l_c}\beta_{k,l,l_c}[n]\mathbf{a}_{\mathrm{M}}\left(\theta_{k,l,l_c}^{\mathrm{M}}, \phi_{k,l,l_c}^{\mathrm{M}}\right) \times \mathbf{a}_{\mathrm{B}}^H\left(\theta_{k,l,l_c}^{\mathrm{B}}, \phi_{k,l,l_c}^{\mathrm{B}}\right),$$

where $\alpha_{k,l,l_c} \sim \mathcal{CN}(0, \frac{N_{\mathrm{B}}N_{\mathrm{M}}}{N_{\mathrm{C}}N_{\mathrm{SC}}})$ is the complex path gain of the l_cth path in the lth cluster with the kth MS; $\beta_{k,l,l_c}[n] = \sum_{d=0}^{D-1} p(dT_{\mathrm{S}} - \tau_{k,l} - \tau_{k,l,l_c})e^{-i\frac{2\pi n}{N}d}$ is the FFT of the samples of the pulse shaping filter function $p(\tau)$, where $\tau_{k,\,l}$ is the delay of the lth cluster with the kth MS, and τ_{k,l,l_c} is the relative delay of the l_cth path in the lth cluster with the kth MS, T_{S} is the sampling period of the pulse shaping filter function, D is the delay spread length, which is also the cyclic prefix length.

In addition, $\mathbf{a}_{\mathrm{B}}(\theta_{pk}^{\mathrm{B}}, \phi_{pk}^{\mathrm{B}}) \in \mathbb{C}^{N_{\mathrm{B}}\times 1}$ is the array steering vector defined as $[\mathbf{a}_{\mathrm{B}}(\theta_{k,l,l_c}^{\mathrm{B}}, \phi_{k,l,l_c}^{\mathrm{B}})]_{n_{\mathrm{B}}} = \frac{1}{\sqrt{N_{\mathrm{B}}}} e^{i\frac{2\pi d_{\mathrm{B}}}{\lambda} n_{\mathrm{Bx}} \cos\phi_{k,l,l_c}^{\mathrm{B}} \sin\theta_{k,l,l_c}^{\mathrm{B}}} \times e^{i\frac{2\pi d_{\mathrm{B}}}{\lambda} n_{\mathrm{By}} \sin\phi_{k,l,l_c}^{\mathrm{B}}}$, where N_{Bx} and N_{By} are the numbers of antennas in the horizontal and vertical directions of the URA at the BS, and satisfy $N_{\mathrm{B}} = N_{\mathrm{Bx}}N_{\mathrm{By}}$; $n_{\mathrm{B}} = n_{\mathrm{By}}N_{\mathrm{Bx}} + n_{\mathrm{Bx}} + 1$, $n_{\mathrm{Bx}} = 0, 1, \cdots, N_{\mathrm{Bx}} - 1, n_{\mathrm{By}} = 0, 1, \cdots, N_{\mathrm{By}} - 1$; d_{B} is the distance between adjacent antenna elements at the BS, λ is the wavelength; $\theta_{k,l,l_c}^{\mathrm{B}}$ and $\phi_{k,l,l_c}^{\mathrm{B}}$ are the corresponding azimuth direction-of-departure (DOD) and elevation DOD at the BS. In addition, $\theta_{k,l,l_c}^{\mathrm{B}} = \theta_{k,l}^{\mathrm{B}} + \vartheta_{k,l,l_c}^{\mathrm{B}}$ and $\phi_{k,l,l_c}^{\mathrm{B}} = \phi_{k,l}^{\mathrm{B}} + \varphi_{k,l,l_c}^{\mathrm{B}}$, where $\theta_{k,l}^{\mathrm{B}}$ and $\phi_{k,l}^{\mathrm{B}}$ are the azimuth DOD and elevation DOD of the lth cluster with the kth MS, $\vartheta_{k,l,l_c}^{\mathrm{B}}$ and $\varphi_{k,l,l_c}^{\mathrm{B}}$ are the relative angle shifts.

Similarly, $\mathbf{a}_{\mathrm{M}}(\theta_{pk}^{\mathrm{M}}, \phi_{pk}^{\mathrm{M}}) \in \mathbb{C}^{N_{\mathrm{M}}\times 1}$ is the array steering vector defined as $[\mathbf{a}_{\mathrm{M}}(\theta_{k,l,l_c}^{\mathrm{M}}, \phi_{k,l,l_c}^{\mathrm{M}})]_{n_{\mathrm{M}}} = \frac{1}{\sqrt{N_{\mathrm{M}}}} e^{i\frac{2\pi d_{\mathrm{M}}}{\lambda} n_{\mathrm{Mx}} \cos\phi_{k,l,l_c}^{\mathrm{M}} \sin\theta_{k,l,l_c}^{\mathrm{M}}} \times e^{i\frac{2\pi d_{\mathrm{M}}}{\lambda} n_{\mathrm{My}} \sin\phi_{k,l,l_c}^{\mathrm{M}}}$, where N_{Mx} and N_{My} are the numbers of antennas in the horizontal and vertical directions of the URA at the MS, and satisfy $N_{\mathrm{M}} = N_{\mathrm{Mx}}N_{\mathrm{My}}$; $n_{\mathrm{M}} = n_{\mathrm{My}}N_{\mathrm{Mx}} + n_{\mathrm{Mx}} + 1$, $n_{\mathrm{Mx}} = 0, 1, \cdots, N_{\mathrm{Mx}} - 1, n_{\mathrm{My}} = 0, 1, \cdots, N_{\mathrm{My}} - 1$; d_{M} is the distance between adjacent antenna elements at the MS; $\theta_{k,l,l_c}^{\mathrm{M}}$ and $\phi_{k,l,l_c}^{\mathrm{M}}$ are the corresponding azimuth direction-of-arrival (DOA) and elevation DOA at the MS. In addition, $\theta_{k,l,l_c}^{\mathrm{M}} = \theta_{k,l}^{\mathrm{M}} + \vartheta_{k,l,l_c}^{\mathrm{M}}$ and $\phi_{k,l,l_c}^{\mathrm{M}} = \phi_{k,l}^{\mathrm{M}} + \varphi_{k,l,l_c}^{\mathrm{M}}$, where $\theta_{k,l}^{\mathrm{M}}$ and $\phi_{k,l}^{\mathrm{M}}$ are the azimuth DOA and elevation DOA of the lth cluster with the kth MS, $\vartheta_{k,l,l_c}^{\mathrm{M}}$ and $\varphi_{k,l,l_c}^{\mathrm{M}}$ are the relative angle shifts. It is assumed that the relative angle shifts, $\vartheta_{k,l,l_c}^{\mathrm{B}}$ and $\varphi_{k,l,l_c}^{\mathrm{B}}$ are beyond the resolution of the BS array; $\vartheta_{k,l,l_c}^{\mathrm{M}}$ and $\varphi_{k,l,l_c}^{\mathrm{M}}$ are beyond the resolution of the MS array.

2.3. Problem formulation

The problem of interest is to design user scheduling strategies that can achieve better tradeoff between capacity and fairness than the existing approaches for mm-wave MIMO systems. It is assumed that perfect channel state information (CSI) is available to the BS in designing user scheduling strategies. It is also assumed that the analog precoders are in the form of array steering vectors with angles corresponding to equally spaced spatial frequencies, i.e., $[\mathbf{F}_t^{\mathrm{R}}]_{:,m} \in \mathcal{B}, \forall m, t$, $\mathbf{w}_{t,k} \in \mathcal{M}, \forall k, t$, where

$$\mathcal{B} = \{\mathbf{a}_{\mathrm{B}}(\vartheta_{\mathrm{B}}, \varphi_{\mathrm{B}}) | \cos\vartheta_{\mathrm{B}}\sin\varphi_{\mathrm{B}} = -1 + 2m_{\mathrm{Bx}}/N_{\mathrm{Bx}}, \sin\vartheta_{\mathrm{B}} = -1 + 2m_{\mathrm{By}}/N_{\mathrm{By}}, m_{\mathrm{Bx}} = 0, 1, \cdots, N_{\mathrm{Bx}} - 1, m_{\mathrm{By}} = 0, 1, \cdots, N_{\mathrm{By}} - 1\}, \quad (2)$$

$$\mathcal{M} = \{\mathbf{a}_{\mathrm{M}}(\vartheta_{\mathrm{M}}, \varphi_{\mathrm{M}}) | \cos\vartheta_{\mathrm{M}}\sin\varphi_{\mathrm{M}} = -1 + 2m_{\mathrm{Mx}}/N_{\mathrm{Mx}}, \sin\vartheta_{\mathrm{M}} = -1 + 2m_{\mathrm{My}}/N_{\mathrm{My}}, m_{\mathrm{Mx}} = 0, 1, \cdots, N_{\mathrm{Mx}} - 1, m_{\mathrm{My}} = 0, 1, \cdots, N_{\mathrm{My}} - 1\}. \quad (3)$$

Additionally, the antenna distances satisfy $d_{\mathrm{B}} = d_{\mathrm{M}} = \lambda/2$.

The tradeoff between efficiency and fairness is important in resource allocation problems, where efficiency is usually the capacity in communication systems. The fairness can be defined in various ways, and the Jain's fairness index is adopted here. In order to achieve the optimal EJT, the scheduling problem can be formulated

into [22]

$$\sigma^{\star} = \max_{\mathbf{r}\in\mathcal{X}} ||\mathbf{r}||_1, \tag{4}$$

where the set $\mathcal{X}$ includes all possible realizations of $\mathbf{r} \in \mathbb{C}^{K\times 1}$ that maximize the Jain's fairness, $\mathbf{r}$ is a vector of the channel capacities. Moreover, $\mathcal{X}$ is defined as

$$\mathcal{X} = \{\mathbf{r}|\mathbf{r} = \arg\max_{\mathbf{r}\in\mathcal{Y}} J(\mathbf{r})\}, \tag{5}$$

where

$$J(\mathbf{r}) = \frac{||\mathbf{r}||_1^2}{K||\mathbf{r}||_2^2} \tag{6}$$

is the Jain's fairness index [16]; $\mathcal{Y}$ is the set of all possible realizations of $\mathbf{r}$ with varying $\mathbf{F}_t^{\mathrm{R}}$, $\mathbf{w}_{t,\,k}$, $\mathbf{F}_t[n]$, and $\mathbf{S}$, i.e.,

$$\begin{aligned}\mathcal{Y} = \{&\mathbf{r}|[\mathbf{F}_t^{\mathrm{R}}]_{:,m} \in \mathcal{B}, \forall m, t,\\ &\mathbf{w}_{t,k} \in \mathcal{M}, \forall k, t,\\ &||\mathbf{F}_t^{\mathrm{R}}\mathbf{F}_t[n]||_{\mathrm{F}}^2 \le P,\\ &[\mathbf{S}]_{t,k} \in \{0,1\}, \forall t, k,\\ &||[\mathbf{S}]_{t,:}||_1 = N_{\mathrm{S}}, \forall t\},\end{aligned} \tag{7}$$

where $t \in \{1, 2, \cdots, T\}$.

In addition, $\mathbf{r}$ is written as

$$[\mathbf{r}]_k = \sum_{t=1}^{T} r_{t,k} \tag{8}$$

being the capacity of the kth MS, where $\mathbf{S} \in \mathbb{R}^{T\times K}$ is the matrix of scheduling variables with $[\mathbf{S}]_{t,k} = 1$ meaning that the kth MS is scheduled at the tth slot and $[\mathbf{S}]_{t,k} = 0$ meaning that the kth MS is not scheduled at the tth slot;

$$\begin{aligned}r_{t,k} = [\mathbf{S}]_{t,k} \sum_{n=1}^{N} \log_2 \Bigg(&1 + |\mathbf{w}_{t,k}^H \mathbf{H}_k[n]\mathbf{F}_t^{\mathrm{R}}[\mathbf{F}_t[n]]_{:,z_{t,k}}|^2\\ &\Big/\Big(1 + \sum_{s\neq z_{t,k}} |\mathbf{w}_{t,k}^H \mathbf{H}_k[n]\mathbf{F}_t^{\mathrm{R}}[\mathbf{F}_t[n]]_{:,s}|^2\Big)\Bigg)\end{aligned} \tag{9}$$

is the capacity of the kth MS at the tth slot; $z_{t,\,k}$ is the index of the element in $\mathbf{s}_t[n]$, $\forall n$ for the kth MS at the tth slot, i.e., $z_{t,k} = \sum_{k'=1}^{k} [\mathbf{S}]_{t,k'}$.

Note that the optimization problem in (5) involves the scheduling of MSs in both the time domain and the spatial domain. More specifically, the design of $\mathbf{S}$ determines the scheduling in the time domain, and the design of $\{\mathbf{F}_t^{\mathrm{R}}\}_{t=1}^{T}, \{\{\mathbf{w}_{t,k}\}_{k=1}^{K}\}_{t=1}^{T}, \{\{\mathbf{F}_t[n]\}_{n=1}^{N}\}_{t=1}^{T}$ determines the scheduling in the spatial domain.

With the signal and channel models as well as the assumptions and the problem, user scheduling will be investigated in the following section.

3. User scheduling in both time and space

As shown earlier, user scheduling for mm-wave MIMO systems is important but is rarely investigated. The existing user scheduling approaches cannot balance performance and fairness to achieve the optimal EJT. In this section, both the spatial resources and the time resources are utilized for scheduling to achieve the optimal EJT.

According to (4) and (5), the user scheduling problem is difficult to solve. This is because the optimization variables, $\{\mathbf{F}_t^{\mathrm{R}}\}_{t=1}^{T}, \{\{\mathbf{w}_{t,k}\}_{k=1}^{K}\}_{t=1}^{T}, \{\{\mathbf{F}_t[n]\}_{n=1}^{N}\}_{t=1}^{T}, \mathbf{S}$, are coupled in the problem. Moreover, the expression of the capacity in (8) is cumbersome and the optimization object in (5) is non-convex. To tackle this problem, we take the following steps.

- First, with the analog beamforming vectors and matrix, $\mathbf{F}_t^{\mathrm{R}}$ and $\{\mathbf{w}_{t,k}\}_{k=1}^{K}$, fixed, the digital beamforming matrices, $\{\mathbf{F}_t[n]\}_{n=1}^{N}$ are designed. This can simplify the problem in (5).
- Then, the time and spatial resources are integrated to make the set $\mathcal{X}$ possess the MTP, and the convex optimization tools are employed to schedule the MSs in both the time and spatial domains.

3.1. Digital beamforming design

According to (1), the upper bound of the system capacity for the nth subcarrier and the tth slot is

$$\begin{aligned}R_t[n] = \log_2 \Big|&\mathbf{I}_{N_{\mathrm{S}}} + \tilde{\mathbf{H}}_t[n]\mathbf{F}_t^{\mathrm{R}}\mathbf{F}_t[n]\\ &\times \mathbf{F}_t[n]^H \mathbf{F}_t^{\mathrm{R}H} \tilde{\mathbf{H}}_t[n]^H\Big|,\end{aligned}$$

where $\tilde{\mathbf{H}}_t[n] \in \mathbb{C}^{N_{\mathrm{S}}\times N_{\mathrm{B}}}$ is the concatenation of the rows $\{\mathbf{w}_{t,k}^H \mathbf{H}_k[n]\}_{k\in\mathcal{U}_t}$, $\mathcal{U}_t$ is the set of the indices of the scheduled MSs at the tth slot and $|\mathcal{U}_t| = N_{\mathrm{S}}$. Without much loss of optimality, the digital beamforming matrix can be designed as

$$\max_{\mathbf{F}_t[n]} R_t[n] \tag{10}$$

$$\text{s.t. } ||\mathbf{F}_t^{\mathrm{R}}\mathbf{F}_t[n]||_{\mathrm{F}}^2 \le P. \tag{11}$$

According to [30,31], the water-filling solution is

$$\mathbf{F}_t[n] = (\mathbf{F}_t^{\mathrm{R}H}\mathbf{F}_t^{\mathrm{R}})^{-\frac{1}{2}} \mathbf{U}_t[n]\mathbf{\Gamma}_t[n], \tag{12}$$

where $\mathbf{U}_t[n] \in \mathbb{C}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ is the right singular matrix of $\tilde{\mathbf{H}}_t[n]\mathbf{F}_t^{\mathrm{R}}(\mathbf{F}_t^{\mathrm{R}H}\mathbf{F}_t^{\mathrm{R}})^{-\frac{1}{2}} \in \mathbb{C}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$, and $\mathbf{\Gamma}_t[n] \in \mathbb{R}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ is a diagonal matrix of the allocated powers to the scheduled MSs at the tth slot and nth subcarrier. More specifically, we denote the singular value decomposition as

$$\tilde{\mathbf{H}}_t[n]\mathbf{F}_t^{\mathrm{R}}(\mathbf{F}_t^{\mathrm{R}H}\mathbf{F}_t^{\mathrm{R}})^{-\frac{1}{2}} = \mathbf{V}_t[n]\mathbf{D}_t[n]\mathbf{U}_t[n]^H, \tag{13}$$

where $\mathbf{V}_t[n] \in \mathbb{C}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ is the left singular matrix, $\mathbf{D}_t[n] \in \mathbb{C}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$ is a diagonal matrix of the singular values. Then, we have (10) simplified as

$$\max_{\mathbf{\Gamma}_t[n]} \log_2 \left|\mathbf{I}_{N_{\mathrm{S}}} + \mathbf{D}_t[n]^2\mathbf{\Gamma}_t[n]^2\right| \tag{14}$$

$$\text{s.t. } ||\mathbf{\Gamma}_t[n]||_{\mathrm{F}}^2 \le P. \tag{15}$$

As can be seen, the water-filling power allocation is

$$\mathbf{\Gamma}_t[n] = \sqrt{\left(\nu\mathbf{I}_{N_{\mathrm{S}}} - \mathbf{D}_t[n]^{-2}\right)^{+}}, \tag{16}$$

where the parameter ν satisfies

$$||\mathbf{\Gamma}_t[n]||_{\mathrm{F}}^2 = P. \tag{17}$$

With the digital beamforming designed, the problem (5) is simplified into

$$\mathcal{X} = \{\mathbf{r}|\mathbf{r} = \arg\max_{\mathbf{r}\in\tilde{\mathcal{Y}}} J(\mathbf{r})\}, \tag{18}$$

where $\tilde{\mathcal{Y}}$ is the set of all possible realizations of $\mathbf{r}$ with varying $\mathbf{F}_t^{\mathrm{R}}$, $\mathbf{w}_{t,\,k}$, and $\mathbf{S}$, i.e.,

$$\begin{aligned}\tilde{\mathcal{Y}} = \{&\mathbf{r}|[\mathbf{F}_t^{\mathrm{R}}]_{:,m} \in \mathcal{B}, \forall m, t,\\ &\mathbf{w}_{t,k} \in \mathcal{M}, \forall k, t,\\ &[\mathbf{S}]_{t,k} \in \{0,1\}, \forall t, k,\\ &||[\mathbf{S}]_{t,:}||_1 = N_{\mathrm{S}}, \forall t\}.\end{aligned} \tag{19}$$

3.2. Monotonic tradeoff property

The optimization problems in (4) and (18) are still hard to solve. The direct way is exhaustive search, but the large cardinalities of $\mathcal{B}$ and $\mathcal{M}$, the large dimension of $\mathbf{S}$, and the coupling of the variables in $\mathbf{r}$ make this method infeasible. The convex optimization is an efficient way but cannot be employed here, because neither the optimization function $J(\mathbf{r})$ nor the optimization set $\tilde{\mathcal{Y}}$ possesses the convexity.

Before simplifying the optimization problems using the MTP, we briefly review the concept of MTP and the related results. The solution of the problem in (4) and (18) can be written as

$$\sigma^\star = \max \mathcal{X}_\sigma^\star, \tag{20}$$

where

$$\mathcal{X}_\sigma^\star = \left\{ \sigma \,|\, \sigma = \arg \max_{\min ||\mathbf{r}||_1 \le \sigma \le \max ||\mathbf{r}||_1} J(\mathbf{r}_\sigma^\star) \right\}, \tag{21}$$

$$\mathbf{r}_\sigma^\star = \arg \min_{||\mathbf{r}||_1 = \sigma, \mathbf{r} \in \tilde{\mathcal{Y}}} ||\mathbf{r}||_2^2. \tag{22}$$

Moreover, the set $\tilde{\mathcal{Y}}$ is said to possess the MTP if $J(\mathbf{r}_\sigma^\star)$ is strictly decreasing with the increase of $||\mathbf{r}_\sigma^\star||_1$ for $||\mathbf{r}_\sigma^\star||_1 > \sigma^\star$, and constant otherwise [22]. In order to use the properties of the MTP, which will be mentioned later, we give the following theorem.

Theorem 1. *The set*

$$\check{\mathcal{Y}} = \left\{ \check{\mathbf{r}} | \check{\mathbf{r}} = \sum_{j=1}^{N_\mathrm{r}} d_j \mathbf{r}_j, d_j \ge 0, \sum_{j=1}^{N_\mathrm{r}} d_j = T \right\}, \tag{23}$$

is convex and possesses the MTP when $d_j, j = 1, 2, \cdots, N_\mathrm{r}$ are continuous, where $\mathbf{r}_j \in \mathbb{C}^{K\times 1}, j = 1, 2, \cdots, N_\mathrm{r}$ are the overall N_r different realizations of $\mathrm{vec}(\{r_{t,k}\}_{k=1}^K)$, and $d_j, j = 1, 2, \cdots, N_\mathrm{r}$ are the corresponding occupation durations for these realizations. Moreover, with $d_j \in \{1, 2, \cdots, T\}, \forall j \in \{1, 2, \cdots, N_\mathrm{r}\}$, $\check{\mathcal{Y}} = \tilde{\mathcal{Y}}$.

Proof. Refer to Appendix A. □

Remark 1. In fact, the coherence time should be shared as discrete slots. However, we can temporarily take the time occupations as continuous and thus use the MTP property and the convexity for optimization. After this, the optimized time occupations can be quantized into discrete values.

Then, we can replace (22) with

$$\mathbf{r}_\sigma^\star = \arg \min_{||\mathbf{r}||_1 = \sigma, \mathbf{r} \in \tilde{\mathcal{Y}}} ||\mathbf{r}||_2^2. \tag{24}$$

Moreover, the user scheduling problem changes into finding the values of $d_j, j = 1, 2, \cdots, N_\mathrm{r}$ in (23) that correspond to $\sigma^\star$ in (20). With MTP being possessed by the problem (20), (21), and (24), we can decrease σ from $\max ||\mathbf{r}||_1$ to $\min ||\mathbf{r}||_1$ and calculate $\mathbf{r}_\sigma^\star$ with (24), the realization of $\mathbf{r}_\sigma^\star$ with which that $J(\mathbf{r}_\sigma^\star)$ stops to increase corresponds to $\sigma^\star$ [22]. Consequently, the user scheduling problem is solved by finding the values of $d_j, j = 1, 2, \cdots, N_\mathrm{r}$ in (23) that correspond to the realization of $\mathbf{r}_\sigma^\star$ with which that $J(\mathbf{r}_\sigma^\star)$ stops to increase.

Since we do not need to calculate (24) for every value of σ, the computational complexity can be reduced. Moreover, as $||\mathbf{r}||_2^2$ in (24) is convex, the problem is more convenient for solving than the original problem in (4) and (18). According to (23), the problem of (24) can be equivalently written as

$$\begin{aligned} \mathbf{x}_\sigma^\star = \arg \min_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{P}^T \mathbf{P} \mathbf{x}, \\ \text{s.t.} \quad & \mathbf{1}_K^T \mathbf{P} \mathbf{x} = \sigma, \\ & \mathbf{1}_{N_\mathrm{r}}^T \mathbf{x} = T, \\ & \mathbf{x} \succeq \mathbf{0}_K, \end{aligned} \tag{25}$$

where $\mathbf{r}_\sigma^\star = \mathbf{P}\mathbf{x}_\sigma^\star$, $\mathbf{x} \in \mathbb{R}^{N_\mathrm{r}\times 1}$ with $[\mathbf{x}]_j = d_j$, $\mathbf{P} \in \mathbb{R}^{K\times N_\mathrm{r}}$ with $[\mathbf{P}]_{:,j} = \mathbf{r}_j$.

It can be seen that this is an inequality constrained convex optimization problem, we can get a suboptimal result of the problem (25) in an iterative manner, with the solution in the sth iteration being

$$\hat{\mathbf{x}}_\sigma^\star(s) = f(\tilde{\mathbf{x}}_\sigma^\star(s)), \tag{26}$$

where $f(\cdot)$ is the function that makes $[f(\tilde{\mathbf{x}}_\sigma^\star(s))]_k = 0, \forall k \in \mathbb{C}_s$ and $[f(\tilde{\mathbf{x}}_\sigma^\star(s))]_k = \xi_s([\tilde{\mathbf{x}}_\sigma^\star(s)]_k - [\hat{\mathbf{x}}_\sigma^\star(s-1)]_k), \forall k \notin \mathbb{C}_s$, $\xi_s = \sum_{\tilde{k}\in\mathbb{C}_s}[f(\tilde{\mathbf{x}}_\sigma^\star(s))]_{\tilde{k}} / (\sum_{k\notin\mathbb{C}_s}[\tilde{\mathbf{x}}_\sigma^\star(s)]_k - [\hat{\mathbf{x}}_\sigma^\star(s-1)]_k)$,

$$\mathbb{C}_s = \{k | [\tilde{\mathbf{x}}_\sigma^\star(s)]_k \le 0\}. \tag{27}$$

Additionally, we have

$$\begin{aligned} \tilde{\mathbf{x}}_\sigma^\star(s) = \arg \min_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{P}^T \mathbf{P} \mathbf{x}, \\ \text{s.t.} \quad & \mathbf{1}_K^T \mathbf{P} \mathbf{x} = \sigma, \\ & \mathbf{1}_{N_\mathrm{r}}^T \mathbf{x} = T, \\ & [\mathbf{x}]_k = [\hat{\mathbf{x}}_\sigma^\star(s-1)]_k, \forall k \in \mathbb{C}_{s-1}, \end{aligned} \tag{28}$$

which is a convex optimization problem and the Newton's method in [36] can be employed to solve.

In accordance with the proposed approach, we demonstrate Algorithm 1 as follows.

Algorithm 1 User Scheduling Based on MTP.

Initialize: $\sigma_{\max} = \max ||\mathbf{r}||_1$, $\sigma_{\min} = \min ||\mathbf{r}||_1$,
$\epsilon_0 = 0.1$, $N_{\mathrm{ni}} = 20$, $\mathbb{C}_0 = \emptyset$
for $n_{\mathrm{ni}} = 0 \to N_{\mathrm{ni}}$ **do**
  $\sigma \leftarrow \sigma_{\max} - n_{\mathrm{ni}}(\sigma_{\max} - \sigma_{\min})/N_{\mathrm{ni}}$
  Find a feasible vector $\hat{\mathbf{x}}_\sigma^\star(0)$ for the constraints in (25)
  $\epsilon_1 \leftarrow 10, \epsilon_2 \leftarrow 100, s \leftarrow 0$
  while $0.5\epsilon_1^2 > \epsilon_0$ and $|\epsilon_1 - \epsilon_2| > 0.1\epsilon_1$
    $s \leftarrow s + 1, \epsilon_2 \leftarrow \epsilon_1$
    Compute Newton step $\mathbf{x}_s$ and decrement $\mathbf{g}_s$ of (28)
    $\epsilon_1 \leftarrow \sqrt{2\mathbf{g}_s^T \mathbf{P}_s^T \mathbf{P}_s \mathbf{g}_s}$
    Choose step size δ_s, calculate $\tilde{\mathbf{x}}_\sigma^\star(s)$
    Calculate (26) and (27)
  end while
  $\mathbf{y}_\sigma^\star(n_{\mathrm{ni}}) \leftarrow \hat{\mathbf{x}}_\sigma^\star(s)$
  $\mathbf{r}_\sigma^\star(n_{\mathrm{ni}}) \leftarrow \mathbf{P}\mathbf{y}_\sigma^\star(n_{\mathrm{ni}})$
  Calculate $J(\mathbf{r}_\sigma^\star(n_{\mathrm{ni}}))$
  if $J(\mathbf{r}_\sigma^\star(n_{\mathrm{ni}})) \le J(\mathbf{r}_\sigma^\star(n_{\mathrm{ni}} - 1))$
    break
  end if
end for
$d_j \leftarrow [\mathbf{y}_\sigma^\star(n_{\mathrm{ni}} - 1)]_j, j = 1, 2, \cdots, N_\mathrm{r}$

Remark 2. N_{ni} is the number of intervals with which we equally divide σ. The Newton step and decrement are $\mathbf{x}_s \in \mathbb{R}^{(N_\mathrm{r} - |\mathbb{C}_{s-1}|)\times 1}$ and $\mathbf{g}_s \in \mathbb{R}^{(N_\mathrm{r} - |\mathbb{C}_{s-1}|)\times 1}$. The intermediate capacity matrix is $\mathbf{P}_s = \mathbf{P}_{:,\{1,2\cdots,N_\mathrm{r}\}\setminus\mathbb{C}_{s-1}} \in \mathbb{R}^{K\times(N_\mathrm{r} - |\mathbb{C}_{s-1}|)}$. This scheduling algorithm utilizes the optimization algorithm in [22] and the Newton's method in [36].

4. Analog beamforming design and user grouping

In this section, the analog beamforming is designed and the users are grouped to further simplify the optimization problem. Although the scheduling can be executed in a low complexity way, the number of choices for optimization, i.e., N_r in (23), may be

very large. For example, with $K = 30$, $N_S = 8$, $N_B = 64$, $N_M = 8$, the proof of Theorem 1 manifests that $N_r = \binom{K}{N_S} \times N_B^{N_S} \times N_M^{N_S} = \binom{30}{8} \times 64^8 \times 8^8$. Apparently, the proposed approach is infeasible for such N_r. In the following, it is demonstrated that the partial choices can be abandoned while keeping the MTP, and the rules for selecting the partial choices are proposed.

Lemma 1. *With part choices of $\mathbf{F}_t^R$ and $\mathbf{w}_{t,k}$ in (19) abandoned, the corresponding set is*

$$\bar{\mathcal{Y}} = \left\{ \bar{\mathbf{r}} | \bar{\mathbf{r}} = \sum_{j=1}^{\bar{N}_r} \bar{d}_j \bar{\mathbf{r}}_j, \bar{d}_j \geq 0, \sum_{j=1}^{\bar{N}_r} \bar{d}_j = T \right\}, \tag{29}$$

where $\bar{\mathbf{r}}_j \in \mathbb{C}^{K\times 1}$, $j = 1, 2, \cdots, \bar{N}_r$ are $\bar{N}_r$ different realizations of $\mathrm{vec}(\{r_{t,k}\}_{k=1}^K)$. In addition, the corresponding occupation durations for these realizations are denoted as $\bar{d}_j$, $j = 1, 2, \cdots, \bar{N}_r$. The set $\bar{\mathcal{Y}}$ is convex and possesses the MTP when $\bar{d}_j$, $j = 1, 2, \cdots, \bar{N}_r$ are continuous.

Proof. The proof is similar to that of Theorem 1 and is omitted for brevity. □

With part of the choices of $\mathbf{F}_t^R$ and $\mathbf{w}_{t,k}$ as well as $[\mathbf{S}]_{t,:}$ in (19) abandoned, the optimization problem can be simplified. More specifically, $\mathbf{r}_\sigma^\star$ in (24) is replaced with

$$\mathbf{r}_\sigma^\star = \arg \min_{||\mathbf{r}||_1 = \sigma, \mathbf{r} \in \bar{\mathcal{Y}}} ||\mathbf{r}||_2^2. \tag{30}$$

Additionally, d_j, $j = 1, 2, \cdots, N_r$ are replaced with $\bar{d}_j$, $j = 1, 2, \cdots, \bar{N}_r$. However, which part of the choices of $\mathbf{F}_t^R$ and $\mathbf{w}_{t,k}$ as well as $[\mathbf{S}]_{t,:}$ is more suitable to be abandoned, i.e., how to design the analog beamforming vectors and group the MSs, is unknown. We will propose a simple way of designing the analog beamforming vectors and the scheduling variables below.

4.1. Cluster separation with beamspace idea

In order to gain high beamforming gain and measure the interference between MSs in a simple way, we resort to the beamspace idea in [37,38]. It is known that the N_B vectors in $\mathcal{B}$, cf. (2), are orthonormal and form the basis of an N_B-dimensional space. Likewise, the N_M vectors in $\mathcal{M}$, cf. (3), are orthonormal and form the basis of an N_M-dimensional space. For convenience, we will term the vectors in $\mathcal{B}$ and these in $\mathcal{M}$ as transmit beams and receive beams, respectively. Correspondingly, the space spanned by the vectors in $\mathcal{B}$ and that spanned by the vectors in $\mathcal{M}$ are named transmit beamspace and receive beamspace, respectively. Note that each vector in $\mathcal{B}$ corresponds to one specific value of the azimuth DOD ϑ_B and one specific value of the elevation DOD φ_B, and each vector in $\mathcal{M}$ corresponds to one specific value of the azimuth DOA ϑ_M and one specific value of the elevation DOA φ_M. As the proposed approach only relies on the orthonormality of the steering vectors of the URA, it is also applicable when other arrays with orthogonal array steering vectors are employed. For example, the array steering vectors of a uniform linear array (ULA) with specifical angles are orthogonal; the uniform cylindrical array in [39] can be turned into a ULA by applying beamforming for each circular array. Thus, the proposed approach is also applicable for these arrays.

Meanwhile, it is known that the correlation of the array steering vector of one transmit path and one transmit beam is dependent on the difference between their spatial frequencies. Accordingly, the transmit beamspace and the receive beamspace can be divided into rectangular areas, and the rectangular areas are denoted as bins. The bin corresponding to m_{Bx} and m_{By} in $\mathcal{B}$ is denoted as the $m_{Bx} + m_{By}N_{Bx}$-th bin in the transmit beamspace, and the bin corresponding to m_{Mx} and m_{My} in $\mathcal{M}$ is denoted as the $m_{Mx} + m_{My}N_{Mx}$th bin in the receive beamspace. Then, the lth cluster of the kth MS is said to be in the $m_{Bx} + m_{By}N_{Bx}$th bin in the transmit beamspace if $|\cos\phi_{k,l}^B \sin\theta_{k,l}^B - (-1 + 2m_{Bx}/N_{Bx})| < 1/N_{Bx}$, $|\sin\phi_{k,l}^B - (-1 + 2m_{By}/N_{By})| < \frac{1}{N_{By}}$. Similarly, the lth cluster of the kth MS is said to be in the $m_{Mx} + m_{My}N_{Mx}$th bin in the receive beamspace if $|\cos\phi_{k,l}^M \sin\theta_{k,l}^M - (-1 + 2m_{Mx}/N_{Mx})| < 1/N_{Mx}$, $|\sin\phi_{k,l}^M - (-1 + 2m_{My}/N_{My})| < \frac{1}{N_{My}}$.

In addition, the indices of the bins of the lth cluster of the kth MS in the transmit beamspace and the receive beamspace are denoted as $b_{k,l}^B$ and $b_{k,l}^M$, respectively. Correspondingly, the sets of these bins are denoted as $\mathcal{B}_k = \{b_{k,l}^B, l = 1, 2, \cdots, N_C\}$, $\mathcal{M}_k = \{b_{k,l}^M, l = 1, 2, \cdots, N_C\}$.

4.2. Analog beamforming and user grouping with beamspace

In order to maximize the beamforming gain, we have the basic rule for choosing the beams.

- Rule 1: The vectors corresponding to $\mathcal{B}_k$ and $\mathcal{M}_k$ should be chosen to form $[\mathbf{F}_t^R]_{:,z_{t,k}}$ and $\mathbf{w}_{t,k}$, respectively.

As can be seen, the beams are steered toward the directions of the lth cluster of the kth MS.

From (13), it can be seen that the analog beamforming vectors should be designed to make sure that they are not the same, otherwise the matrix inversion $(\mathbf{F}_t^{R^H}\mathbf{F}_t^R)^{-1}$ does not exist. In order to avoid this problem, another rule for selecting the precoding vectors is proposed.

- Rule 2: If one bin in the transmit beamspace is shared by $K' > 1$ MSs that are scheduled at the t-th slot, i.e., $\mathcal{B}_{k_1} \cap \mathcal{B}_{k_2} \cdots \cap \mathcal{B}_{k_{K'}} = b_{k_1,l}^B$, then the vector corresponding to $b_{k_1,l}^B$ in $\mathcal{B}$ can be chosen to form at most one column of $\mathbf{F}_t^R$.

According to the original problem in (5), it is known that only the elements with large Jain's fairness index in $\mathcal{Y}$ may be the optimal scheduling solution. The definition of the elements in $\mathbf{r}$, i.e., $r_{t,k}$, in (9) manifests that the lower the interference, the closer the elements in $\mathbf{r}$; consequently, the larger the Jain's fairness index. Therefore, we can pick out the elements in $\bar{\mathcal{Y}}$ with low interference to form $\bar{\mathcal{Y}}$. Correspondingly, we have the following rule.

- Rule 3: If one bin in the receive beamspace is shared by C' clusters of the k_1th MS, i.e., $b_{k_1,l_1}^M = b_{k_1,l_2}^M \cdots = b_{k_1,l_{C'}}^M$, and the vector corresponding to one of the bins in the transmit beamspace that correspond to these clusters is chosen to form the precoding vector for this MS, i.e., the vector corresponding to b_{k_1,l_1}^B forms $[\mathbf{F}_t^R]_{:,z_{t,k}}$; then, vectors corresponding to other beams in the transmit beamspace, i.e., $b_{k_1,l}^B, l \in \{l_2, \cdots, l_{C'}\}$, cannot be chosen to form the precoding vector for any MS.

In order to further decrease the number of selected choices, we choose one MS as less times as possible. Since the digital beamforming favors the dominant user, i.e., the user that corresponds to the first row of $\tilde{\mathbf{H}}_t[n]$, each user should have the opportunity of being the dominant user. This restriction is written in a general form as the following rule.

- Rule 4: Denote the set of the selected realizations of $[\mathbf{S}]_{t,:}$ as $\mathcal{S} = \{\mathbf{s}_{\bar{n}}\}_{\bar{n}=1}^{\bar{N}_r}$, where $\mathbf{s}_{\bar{n}} \in \mathbb{R}^{K\times 1}$. Then, $\sum_{\bar{n}=1}^{\bar{N}_r} [\mathbf{s}_{\bar{n}}]_k \geq 1, \forall k$. Moreover, for $\{\bar{n} | [\mathbf{s}_{\bar{n}}]_k = 1\}$, the kth MS should be the dominant user for only one element in this set.

Based on these rules, we have $\bar{N}_r = K$, which makes the proposed Algorithm 1 feasible. Additionally, these rules guarantee a good system performance. Consequently, we propose the following Algorithm 2 for choosing the beamforming vectors and grouping the MSs, which correspond to $\mathbf{F}_t^R$, $\mathbf{w}_{t,k}$, and $[\mathbf{S}]_{t,:}$ in (19).

Algorithm 2 Selection of the beamforming vectors and user grouping.

Initialize: Groups of indices of selected MSs $\mathcal{G}_k = \emptyset, \forall k \in \{1, 2, \cdots, K\}$
1: **for** $k = 1 \rightarrow K$ **do**
2: $\mathcal{G}_k \leftarrow \mathcal{G}_k \bigcup k$, $l^{\star}_{1,k} \leftarrow 1$, $j \leftarrow k$
3: **while** $|\mathcal{G}_k| < N_{\mathrm{S}}$
4: $j \leftarrow j + 1 - \lfloor (j + 0.1)/K \rfloor K$
5: **for** $l = 1 \rightarrow N_{\mathrm{C}}$
6: **if** $b^{\mathrm{B}}_{j,l}$ and $b^{\mathrm{M}}_{j,l}$ satisfy Rule 2 and Rule 3
7: $\mathcal{G}_k \leftarrow \mathcal{G}_k \bigcup j$, $l^{\star}_{j,k} \leftarrow l$
8: **break**
9: **end**
10: **end for**
11: **end while**
12: Form $\mathbf{F}^{\mathrm{R}}_t$ with elements corresponding to $l^{\star}_{j,k}$, $j \in \mathcal{G}_k$ in $\mathcal{B}$, form $\mathbf{w}_{t,k}$ with elements corresponding to $l^{\star}_{j,k}$ in $\mathcal{M}$, form $[\mathbf{S}]_{t,:}$ according to $\mathcal{G}_k$
13: **end for**

Remark 3. For steps 2 and 7, we denote the selected cluster for the jth MS in the kth group as $l^{\star}_{j,k}$. Moreover, $\mathbf{F}^{\mathrm{R}}_t$, $\mathbf{w}_{t,k}$, and $[\mathbf{S}]_{t,:}$ generated in step 12 each time correspond to one choice of the user scheduling and the beam selection, and also correspond to one realization of $\bar{\mathbf{r}}_j$ given in (29). In other words, with the selected beams and the grouped MSs, we can calculate each $\bar{\mathbf{r}}_j$ according to (9).

4.3. Computational complexity analysis and comparison

With the traditional approaches employed, their computational complexities can be analyzed and are listed in Table 1. The big O notation is used to denote the number of multiplicities required as the variables grow. The details are as follows. The greedy approach calculates the capacity when each unscheduled MS is scheduled in the current scheduling step, and then chooses the MS that results into the largest capacity. Thus, (9) and (12) should be calculated each time. By taking the computations in these equations into consideration, it can be found that the computational complexity for choosing the first MS is $O(NN_{\mathrm{B}}KN_{\mathrm{M}})$, the computational complexity for choosing the N_{R}th MS is $O(NN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$. Thus, the computational complexity of the greedy approach is at least $O(NN_{\mathrm{B}}(KN_{\mathrm{M}} + N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}})))$.

The max-min scheduling approach schedules the N_{R} MSs with the smallest average sum rate in each time slot. Thus, (9) and (12) should be calculated each time slot. The number of time slots is T. Thus, the computational complexity is $O(TNN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$.

The PF scheduling approach schedules the N_{R} MSs with the largest ratio of the current rate to the average sum rate in each time slot. Thus, the computational complexity of PF is the same as that of the max-min.

Table 1
Comparison of computational complexities.

Methods	Computational Complexities
Greedy	$O(NN_{\mathrm{B}}(KN_{\mathrm{M}} + N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}})))$
Max-min	$O(TNN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$
PF	$O(TNN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$
RR	$O(1)$
OR in [27,40]	$O(1)$
Proposed	$O(KNN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$

Table 2
Simulation parameters.

Parameter	Value	Parameter	Value
N_{Bx}	16	N_{By}	4
N_{Mx}	4	N_{My}	2
N_{R}	8	K	30
N_{C}	6	N_{SC}	5
D	8	N	32
T	50	P/σ^2	10

The RR approach schedules MSs in predefined order, which means it does not need any computation. The computational complexity is $O(1)$.

The proposed approach calculates $\bar{\mathbf{r}}_j$, $j = 1, 2, \cdots, K$ in (29), which means (9) and (12) should be calculated for K times. Thus, the computational complexity is $O(KNN_{\mathrm{B}}N_{\mathrm{R}}(N_{\mathrm{M}} + N_{\mathrm{R}}))$.

With typical system parameters given in Table 2, it can be seen that the computational complexity of the proposed approach is close to that of the greedy approach, the RR approach, the max-min approach. What's more, two state-of-the-art user scheduling approaches are also compared here. An orthogonal criterion based user scheduling approach is proposed in [40], which aims at orthogonalizing the steering vectors for the DOAs of the scheduling users. Another orthogonal criterion for user scheduling is proposed in [27], which schedules users with the least overlapping with the already scheduled users in the beamspace. As these two approaches only need to calculate the DOAs, the computational complexity of each one is $O(1)$. In Table 1 and in the following, the two approaches are denoted as OR. These OR approaches are similar to the proposed approach in the sense that they schedule MSs according to the DOAs. The major difference is that the OR approaches aims at maximizing the capacity, but the proposed approach tries to balance capacity and fairness.

5. Numerical results

In this section, numerical results are provided to compare the tradeoff of the proposed approach and that of other existing approaches. The simulation parameters are given in Table 2. The pulse shaping filter is [34]

$$p(t) = \begin{cases} \frac{\pi}{4}\mathrm{sinc}(\frac{1}{2\beta}), & t = \pm\frac{T_{\mathrm{S}}}{2\beta}, \\ \mathrm{sinc}(\frac{t}{T_{\mathrm{S}}})\frac{\cos(\frac{\pi\beta t}{T_{\mathrm{S}}})}{1-(\frac{2\beta t}{T_{\mathrm{S}}})^2}, & \text{otherwise}, \end{cases} \tag{31}$$

where $\beta = 1$. The cluster delay, $\tau_{k,l}/T_{\mathrm{S}}$ is uniformly distributed in [0, D]. The relative delay, $\tau_{k,l,l_c}/T_{\mathrm{S}}$ is uniformly distributed in [0, 0.05D]. For the cluster angles, $\theta^{\mathrm{B}}_{k,l}$ is uniformly distributed in $[-\pi/6, \pi/6]$, $\phi^{\mathrm{B}}_{k,l}$ is uniformly distributed in [0, $\pi/2$], $\theta^{\mathrm{M}}_{k,l}$ is uniformly distributed in [0, 2π], $\phi^{\mathrm{M}}_{k,l}$ is uniformly distributed in [0, $\pi/2$]. The relative angle shifts, $\vartheta^{\mathrm{B}}_{k,l,l_c}$, $\varphi^{\mathrm{B}}_{k,l,l_c}$, $\vartheta^{\mathrm{M}}_{k,l,l_c}$, $\varphi^{\mathrm{M}}_{k,l,l_c}$, are uniformly distributed in [0, 0.05π].

In Fig. 2, the capacity and the Jain's fairness index versus the number of the SNR P/σ^2 are demonstrated. It can be seen that the proposed approach can achieve the same fairness as the max-min approach, and achieves the highest fairness among all the approaches. Meanwhile, the capacity of this approach is higher than the max-min and the PF approach. The capacity of the proposed approach is lower than the greedy approach and the OR approaches in [27,40]. But the fairness of the proposed approach is much higher than that of these approaches. These results verify that the proposed approach approximately achieves the EJT, as its fairness is the highest and its capacity is higher than that of approaches with similar fairness.

In Fig. 3, the capacity and the Jain's fairness index versus the number of the MSs K are demonstrated. It can be seen that the

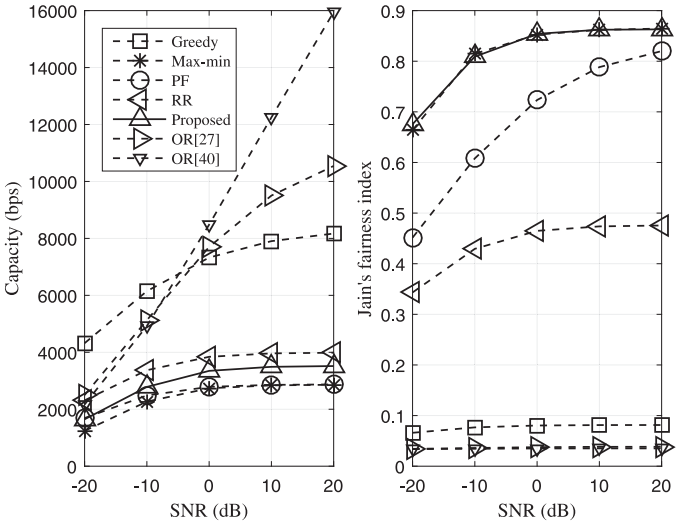

Fig. 2. The left figure demonstrates the capacities versus the SNR P/σ^2. The right figure demonstrates the Jain's fairness index.

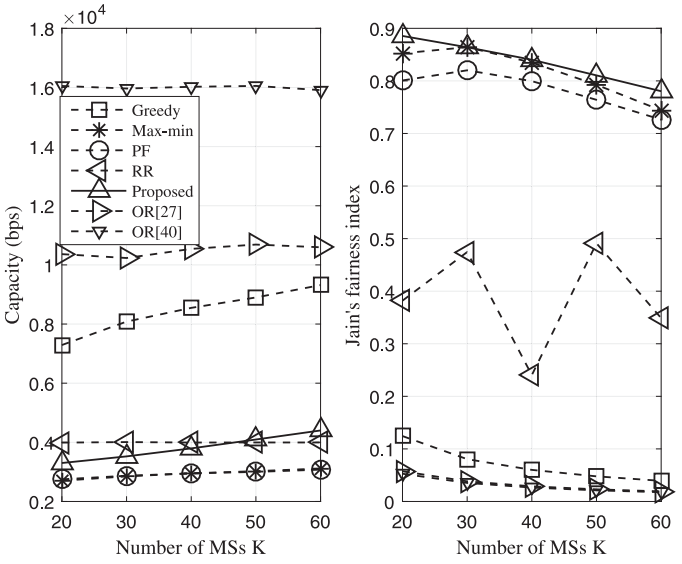

Fig. 3. The left figure demonstrates the capacities versus the number of MSs K. The right figure demonstrates the Jain's fairness index.

proposed approach is still of the highest fairness. Additionally, the fairness of the proposed approach surpasses that of the max-min for some values of K. As K grows, the capacity of the proposed approach increases faster than that of other approaches and begins to surpass other approaches except the greedy approach when $K = 50$.

In Fig. 4, the capacity and the Jain's fairness index versus the number of the BS antennas N_{B} are demonstrated. As N_{Bx} changes, the number of the vertical BS antennas N_{By} changes in accordance. Here, the utilized combinations of N_{Bx} and N_{By} are (16, 4), (18, 6), (20, 8), (22, 10), (24, 12). As can be seen, the fairness of the proposed approach is still higher than that of other approaches. Meanwhile, the capacity and the fairness are almost invariant with the increase of N_{B} except that of the greedy approach, which means the number of BS antennas needs not to be very large for the proposed approach.

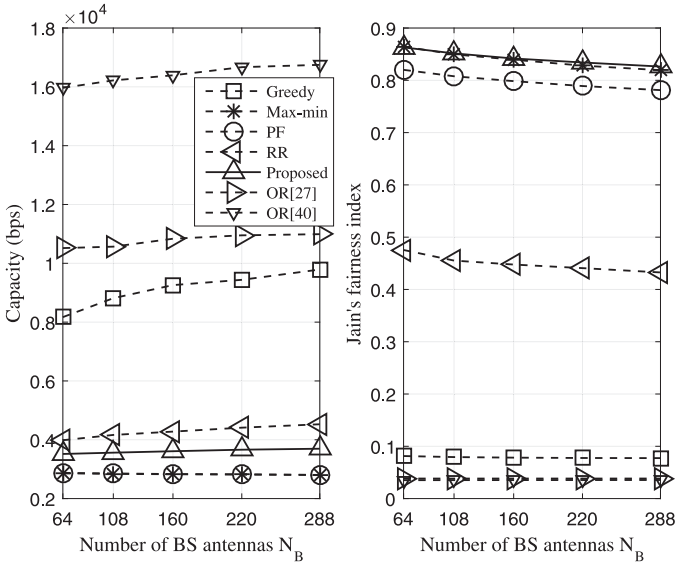

Fig. 4. The left figure demonstrates the capacities versus the number of BS antennas N_{B}. The right figure demonstrates the Jain's fairness index.

6. Conclusions

In this paper, user scheduling that balances between capacity and Jain's fairness index in mm-wave MIMO systems is investigated. The digital precoder is first derived, which simplifies the problem into a joint selection of the analog precoder, analog combiner, and the user grouping. Based on the MTP possessed by the investigated problem, traditional convex optimization algorithms can be employed to solve the problem. Then, the sets of the analog precoder, analog combiner, and user group are condensed by abandoning unnecessary parts. Finally, the computational complexity is analyzed and the performance of the proposed approach is compared with other approaches by the numerical results.

Acknowledgments

This research was supported by Project 61601152 supported by National Natural Science Foundation of China.

Appendix A. Proof of Theorem 1

Before the proof, we give a lemma first.

Lemma 2. *The set* $\tilde{\mathcal{Y}}$ *is convex. With* $d_j \in \{1, 2, \cdots, T\}$, $\forall j \in \{1, 2, \cdots, N_{\mathrm{r}}\}$, $\tilde{\mathcal{Y}} = \tilde{\mathcal{Y}}$.

Proof. From (9), it can be seen that the values of $\{r_{t,k}\}_{k=1}^{K}$ depend on the choice of $\mathbf{F}_t^{\mathrm{R}}$, $\{\mathbf{w}_{t,k}\}_{k=1,[\mathbf{S}]_{t,k}\neq 0}^{K}$, and $[\mathbf{S}]_{t,:}$. According to the last two restrictions in (19), there are $\binom{K}{N_{\mathrm{S}}}$ different realizations of $[\mathbf{S}]_{t,:}$, and the set that contains these realizations is denoted as $\mathcal{A}_{\mathrm{S}}$. According to the first restriction in (19), $[\mathbf{F}_t^{\mathrm{R}}]_{:,m}$ has N_{B} different realizations. Thus, $\mathbf{F}_t^{\mathrm{R}}$ has $N_{\mathrm{B}}^{N_{\mathrm{S}}}$ different realizations, and the set that contains these realizations is denoted as $\mathcal{A}_{\mathrm{B}}$. According to the second restriction in (19), $\mathbf{w}_{t,k}$ has N_{M} different realizations. Thus, $\{\mathbf{w}_{t,k}\}_{k=1,[\mathbf{S}]_{t,k}\neq 0}^{K}$ has $N_{\mathrm{M}}^{N_{\mathrm{S}}}$ different realizations, and the set that contains these realizations is denoted as $\mathcal{A}_{\mathrm{M}}$. As a result, there are $|\mathcal{A}_{\mathrm{S}}||\mathcal{A}_{\mathrm{M}}||\mathcal{A}_{\mathrm{B}}|$ choices of $\mathbf{F}_t^{\mathrm{R}}$, $\{\mathbf{w}_{t,k}\}_{k=1,[\mathbf{S}]_{t,k}\neq 0}^{K}$, and $[\mathbf{S}]_{t,:}$, and (19) can be equivalently expressed as $\tilde{\mathcal{Y}} = \{\mathbf{r}|\mathbf{F}_t^{\mathrm{R}} \in \mathcal{A}_{\mathrm{B}}, \forall t, \{\mathbf{w}_{t,k}\}_{k=1,[\mathbf{S}]_{t,k}\neq 0}^{K} \in \mathcal{A}_{\mathrm{M}}, \forall t, [\mathbf{S}]_{t,:} \in \mathcal{A}_{\mathrm{S}}, \forall t\}$.

Note this also means that there are $N_{\mathrm{r}} \leq |\mathcal{A}_{\mathrm{S}}||\mathcal{A}_{\mathrm{M}}||\mathcal{A}_{\mathrm{B}}|$ different realizations of $\mathrm{vec}(\{r_{t,k}\}_{k=1}^{K})$, and these realizations are denoted as

$\mathbf{r}_j \in \mathbb{C}^{K\times 1}, j = 1, 2, \cdots, N_\mathrm{r}$. Moreover, there is an occupation duration for each realization $\mathbf{r}_j$ in the coherence interval T, which is denoted as d_j. As the coherence time is T, we have $\sum_{j=1}^{N_\mathrm{r}} d_j = T$. Then, according to the definition of $\mathbf{r}$ in (8), (19) can be re-written as $\tilde{\mathcal{Y}} = \{\mathbf{r} | \mathbf{r} = \sum_{j=1}^{N_\mathrm{r}} d_j \mathbf{r}_j, d_j \in \{1, 2, \cdots, T\}, \sum_{j=1}^{N_\mathrm{r}} d_j = T\}$. By changing discrete $d_j, j = 1, 2, \cdots, N_\mathrm{r}$ into continuous $\check{d}_j, j = 1, 2, \cdots, N_\mathrm{r}$, $\tilde{\mathcal{Y}}$ is changed into $\check{\mathcal{Y}}$. For $\check{\mathcal{Y}}$, we have $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \check{\mathcal{Y}}, \forall \mathbf{x}_1, \mathbf{x}_2 \in \check{\mathcal{Y}}, 0 \le \alpha \le 1$. Thus, the set $\check{\mathcal{Y}}$ is convex. □

Then, we create a new set

$$\check{\mathcal{Y}}_\mathrm{a} = \left\{ \check{\mathbf{r}}_\mathrm{a} | \check{\mathbf{r}}_\mathrm{a} = \sum_{j=1}^{N_\mathrm{r}+1} d_j \mathbf{r}_j, d_j \ge 0, \sum_{j=1}^{N_\mathrm{r}+1} d_j = T \right\}, \tag{32}$$

where $\mathbf{r}_{N_\mathrm{r}+1} = \mathbf{0}_K$. With $d_{N_\mathrm{r}+1} = T$, we have $\check{\mathbf{r}}_\mathrm{a} = \mathbf{0}_K$; thus, we have $\mathbf{0}_K \in \check{\mathcal{Y}}_\mathrm{a}$. Since $\mathbf{r}_j \succeq \mathbf{0}_K$, there is $\check{\mathbf{r}}_\mathrm{a} \succeq \mathbf{0}_K, \forall \check{\mathbf{r}}_\mathrm{a} \in \check{\mathcal{Y}}_\mathrm{a}$. Based on Theorem 1 in [22], we conclude that the set $\check{\mathcal{Y}}_\mathrm{a}$ possesses the MTP.

According to the definition of $\check{\mathcal{Y}}$ in (23), we have $\check{\mathcal{Y}} \cup \check{\mathcal{Y}}_\mathrm{s} = \check{\mathcal{Y}}_\mathrm{a}$, where $\check{\mathcal{Y}}_\mathrm{s} = \{\check{\mathbf{r}}_\mathrm{s} | \check{\mathbf{r}}_\mathrm{s} = \sum_{j=1}^{N_\mathrm{r}+1} d_j \mathbf{r}_j, d_j \ge 0, d_{N_\mathrm{r}+1} > 0, \sum_{j=1}^{N_\mathrm{r}+1} d_j = T\}$. Meanwhile, for any $\check{\mathbf{r}}_\mathrm{s} \in \check{\mathcal{Y}}_\mathrm{s}$, there is a counterpart $\check{\mathbf{r}} \in \check{\mathcal{Y}}$, where $\check{\mathbf{r}} = \sum_{j=1}^{N_\mathrm{r}} (d_j + d_j d_{N_\mathrm{r}+1} / \sum_{jj=1}^{N_\mathrm{r}} d_{jj}) \mathbf{r}_j$. It can be seen that $J(\check{\mathbf{r}}) = J(\check{\mathbf{r}}_\mathrm{s})$ and $||\check{\mathbf{r}}_\mathrm{s}||_1 < ||\check{\mathbf{r}}||_1$. As $\check{\mathcal{Y}}_\mathrm{a}$ possesses the MTP, $J(\mathbf{r}^\star_\gamma)$ is strictly decreasing with the increase of $||\mathbf{r}^\star_\gamma||_1$ for $||\mathbf{r}^\star_\gamma||_1 > \gamma^\star$, where $\mathbf{r}^\star_\gamma = \arg\min_{||\check{\mathbf{r}}_\mathrm{a}||_1 = \gamma, \check{\mathbf{r}}_\mathrm{a} \in \check{\mathcal{Y}}_\mathrm{a}} ||\mathbf{r}||_2^2$. Hence, there are two cases for discussion. 1) $||\check{\mathbf{r}}_\mathrm{s}||_1 > \gamma^\star$, then we have $J(\mathbf{r}^\star_{||\check{\mathbf{r}}_\mathrm{s}||_1}) > J(\mathbf{r}^\star_{||\check{\mathbf{r}}||_1}) \ge J(\check{\mathbf{r}})$, which means $\mathbf{r}^\star_{||\check{\mathbf{r}}_\mathrm{s}||_1} \ne \check{\mathbf{r}}_\mathrm{s}$. 2) $||\check{\mathbf{r}}_\mathrm{s}||_1 \le \gamma^\star$. In conclusion, $\check{\mathbf{r}}_\mathrm{s}$ has no effect on the MTP of $\check{\mathcal{Y}}_\mathrm{a}$. Consequently, $\check{\mathcal{Y}}$ possesses the MTP.

References

[1] Z. Pi, F. Khan, An introduction to millimeter-wave mobile broadband systems, IEEE Commun. Mag. 49 (6) (2011) 101–107.
[2] T.S. Rappaport, S. Sun, R. Mayzus, et al., Millimeter wave mobile communications for 5g cellular: it will work!, IEEE Access 1 (2013) 335–349.
[3] A.L. Swindlehurst, E. Ayanoglu, P. Heydari, et al., Millimeter-wave massive MIMO: the next wireless revolution? IEEE Commun. Mag. 52 (9) (2014) 56–62.
[4] J. Yuan, S. Jin, W. Xu, et al., User-centric networking for dense c-RANS: high-SNR capacity analysis and antenna selection, IEEE Trans. Commun. 65 (11) (2017) 5067–5080.
[5] W. Tan, M. Matthaiou, S. Jin, et al., Spectral efficiency of DFT-based processing hybrid architectures in massive MIMO, IEEE Wireless Commun. Lett. 6 (5) (2017) 586–589.
[6] W. Tan, D. Xie, J. Xia, et al., Spectral and energy efficiency of massive MIMO for hybrid architectures based on phase shifters, IEEE Access 6 (2018) 11751–11759.
[7] Z. Shen, R. Chen, J.G. Andrews, et al., Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization, IEEE Trans. Signal Process. 54 (9) (2006) 3658–3663.
[8] L.N. Tran, M. Bengtsson, B. Ottersten, Iterative precoder design and user scheduling for block-diagonalized systems, IEEE Trans. Signal Process. 60 (7) (2012) 3726–3739.
[9] Z. Tu, R.S. Blum, Multiuser diversity for a dirty paper approach, IEEE Commun. Lett. 7 (8) (2003) 370–372.
[10] X. Zhang, J. Lee, Low complexity MIMO scheduling with channel decomposition using capacity upperbound, IEEE Trans. Commun. 56 (6) (2008) 871–876.
[11] A. Razi, D.J. Ryan, I.B. Collings, et al., Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding, IEEE Trans. Wireless Commun. 9 (1) (2010) 356–365.
[12] T. Yoo, N. Jindal, A. Goldsmith, Multi-antenna downlink channels with limited feedback and user selection, IEEE J. Sel. Areas Commun. 25 (7) (2007) 1478–1491.
[13] T. Yoo, A. Goldsmith, On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming, IEEE J. Sel. Areas Commun. 24 (3) (2006) 528–541.
[14] A. Silberschatz, P.B. Galvin, G. Gagne, Operating system concepts, John Wiley & Sons, Hoboken, USA, 2013.
[15] C. Simon, G. Leus, Round-robin scheduling for orthogonal beamforming with limited feedback, IEEE Trans. Wireless Commun. 10 (8) (2011) 2486–2496.
[16] R. Jain, D. Chiu, W. Hawe, A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems, in: M.A. Huston (Ed.), Digital Equipment Corporation, 1984. Tech. Rep. DEC-TR-301.
[17] F.P. Kelly, Charging and rate control for elastic traffic, Eur. Trans. Telecommun. 8 (1) (1997) 33–37.
[18] A. Jalali, R. Padovani, R. Pankaj, Data Throughput of CDMA-HDR a High Efficiency-high Data Rate Personal Communication Wireless System, in: Proc. IEEE 51st Veh. Technol. Conf., 2000, pp. 1854–1858.
[19] S. Bai, W. Zhang, Y. Liu, et al., Max-min Fair Scheduling in OFDMA-Based Multi-hop wiMAX Mesh Networks, in: Proc. IEEE Int. Conf. Commun., 2011, pp. 1–5.
[20] J. Tang, G. Xue, W. Zhang, Maximum Throughput and Fair Bandwidth Allocation in Multi-channel Wireless Mesh Networks, in: Proc. IEEE Int. Conf. Computer and Commun., 2006, pp. 1–10.
[21] B. Song, Y.H. Lin, R.L. Cruz, Weighted max-min fair beamforming, power control, and scheduling for a MISO downlink, IEEE Trans. Wireless Commun. 7 (2) (2008) 464–469.
[22] A.B. Sediq, R.H. Gohary, R. Schoenen, et al., Optimal Tradeoff between Sum-rate Efficiency and Jain's Fairness Index in Resource Allocation, in: IEEE Trans. Wireless Commun., 12, 2013, pp. 3496–3509.
[23] L.X. Cai, L. Cai, X. Shen, et al., REX: A randomized exclusive region based scheduling scheme for mmwave WPANs with directional antenna, IEEE Trans. Wireless Commun. 9 (1) (2010) 113–121.
[24] Z. He, S. Mao, S. Kompella, et al., On link scheduling in dual-hop 60-GHz mmwave networks, IEEE Trans. Veh. Technol. 66 (12) (2017) 11180–11192.
[25] Y. Niu, C. Gao, Y. Li, et al., Energy-efficient scheduling for mmwave backhauling of small cells in heterogeneous cellular networks, IEEE Trans. Veh. Technol. 66 (3) (2017) 2674–2687.
[26] J. García-Rois, F. Gómez-Cuba, M.R. Akdeniz, et al., On the analysis of scheduling in dynamic duplex multihop mmwave cellular systems, IEEE Trans. Wireless Commun. 14 (11) (2015) 6028–6042.
[27] A. Adhikary, E.A. Safadi, M.K. Samimi, et al., Joint spatial division and multiplexing for mm-wave channels, IEEE J. Sel. Areas Commun. 32 (6) (2014) 1239–1255.
[28] R.W. Heath Jr., N. González-Prelcic, S. Rangan, et al., An overview of signal processing techniques for millimeter wave MIMO systems, IEEE J. Sel. Top. Signal Process. 10 (3) (2016) 436–453.
[29] K. Venugopal, A. Alkhateeb, N.G. Prelcic, et al., Channel estimation for hybrid architecture-based wideband millimeter wave systems, IEEE J. Sel. Areas Commun. 35 (9) (2017) 1996–2009.
[30] F. Sohrabi, W. Yu, Hybrid digital and analog beamforming design for large-scale antenna arrays, IEEE J. Sel. Top. Signal Process. 10 (3) (2016) 501–513.
[31] F. Sohrabi, W. Yu, Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays, IEEE J. Sel. Areas Commun. 35 (7) (2017) 1432–1443.
[32] S. Park, A. Alkhateeb, R.W. Heath Jr., Dynamic subarrays for hybrid precoding in wideband mmwave MIMO systems, IEEE Trans. Wireless Commun. 16 (5) (2017) 2907–2920.
[33] Z. Gao, C. Hu, L. Dai, et al., Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels, IEEE Commun. Lett. 20 (6) (2016) 1259–1262.
[34] A. Alkhateeb, R.W. Heath Jr., Frequency selective hybrid precoding for limited feedback millimeter wave systems, IEEE Trans. Commun. 64 (5) (2016) 1801–1818.
[35] K. Venugopal, González-Prelcic, R.W. Heath Jr., Optimality of frequency flat precoding in frequency selective millimeter wave channels, IEEE Wireless Commun. Lett. 6 (3) (2017) 330–333.
[36] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, Cambridge, UK, 2004.
[37] J. Brady, N. Behdad, A.M. Sayeed, Beamspace MIMO for millimeter-wave communications: system architecture, modeling, analysis, and measurements, IEEE Trans. Antennas Propag. 61 (7) (2013) 3814–3827.
[38] J. Brady, A. Sayeed, Beamspace MU-MIMO for High-density Gigabit Small Cell Access at Millimeter-wave Frequencies, in: Proc. IEEE 15th Int. Workshop Signal Process. Advances Wireless Commun, 2014, pp. 80–84.
[39] W. Tan, X. Li, D. Xie, et al., On the performance of three dimensional antenna arrays in millimeter wave propagation environments, IET Commun. 12 (17) (2018) 1743–1750.
[40] W. Tan, S. Jin, C.K. Wen, et al., Spectral efficiency of multi-user millimeter wave systems under single path with uniform rectangular arrays, EURASIP J. Wireless Commun. Networking 181 (2017) 458–472.