Statistical CSIT Aided User Scheduling for Broadcast MU-MISO System

Qi Cao, Yanjing Sun, *Member, IEEE,* Qiang Ni, *Senior Member, IEEE,* Song Li, *Member, IEEE,* and Zefu Tan

Abstract—Recent studies show that the statistical channel state information (SCSI) helps to largely increase the capacity of communication systems when the instantaneous perfect CSI (IPCSI) is unavailable. In this paper, we consider multi-user multiple-input-single-output (MU-MISO) broadcast channels where the transmitter has the knowledge of SCSI. The major issue concerned in our work is to improve the average group-rate of the whole system by scheduling users over different time slots. With SCSI at the transmitter side, we are able to precode signals and hence compute the theoretical achievable group-rate of arbitrary user groups. Based on the group-rates, we propose tier-2 Munkres user scheduling algorithm (T2-MUSA) which leads to higher average group-rate than existing algorithms with generally better fairness. The optimality of the proposed algorithm in energy-fair user scheduling space is proved and we derive a lower bound of a special case to verify the validity of our simulations. In addition, many conventional user scheduling algorithms maintain queue stability by solving a weighted sum-rate (WSR) problem, using queue lengths to represent weight coefficients. Inspired by T2-MUSA we propose a QoS-based Munkres user scheduling algorithm (QB-MUSA) aimed at stabilizing queue lengths and maximizing throughput. In results, we show that QB-MUSA exhibits higher throughput than the conventional weighted sum-rate (WSR) based algorithm.

Index Terms—MISO, Multi-user scheduling, Statistical CSI, Munkres algorithm.

I. INTRODUCTION

MUltiple-input multiple-output (MIMO) system as a hot topic in wireless communication is able to bring substantial improvement to system capacity and has attracted considerable attention from relevant researchers. Early MIMO studies mainly focused on single-user (SU) scenarios [1], [2], while in some later works the focus has transitioned to multi-user (MU) systems [3]–[5] which could realize higher data-rates by utilizing space-division multiple-access (SDMA) and multi-user diversity. In MU-MISO broadcast channel (BC), many works are based on a crucial condition that instantaneous and perfect channel state information (IPCSI) is available at

Qi Cao, Yanjing Sun, Song Li and Zefu Tan are with School of Information and Control Engineering, China University of Mining Technology, Xuzhou, 221116, China (emails: qcao, yjsun, lisong@cumt.edu.cn; tanzefu@163.com).

Qiang Ni is with the School of Computing and Communications, Info-Lab21, Lancaster University, UK, (email: q.ni@lancaster.ac.uk).

the transmitter side [6]–[8]. Whereas, if IPCSI is unavailable, these transmission schemes degrade rapidly or even fail to work [9]. Practically IPCSI is not feasible especially in frequency-division duplex (FDD) channels. Generally speaking, non-IPCSI can be divided into imperfect CSI [9]–[12] and delayed CSI [13]–[15]. In particular, the CSI on the transmitter side (CSIT) is usually obtained via feedback links, however in BC system, the finite-rate uplink cannot accurately describe the exact CSI. Therefore, the transmitter side only acquires an estimated version of CSI with inevitable estimation errors. On the other hand, the feedback delay can possibly render the CSIT totally uncorrelated with IPCSI.

Apart from the CSI categories mentioned above, experimental results show that statistical CSI (SCSI) as a feasible and slowly varied measurement can be easily and accurately obtained by the transmitter side [16], [17]. This measurement is usually named after spatial correlation, antenna correlation or joint correlation in initial works, and it turns out to be very helpful in precoder design, transmission scheme scheduling and thereby capacity estimation [3]–[5], [15], [18]. In SU-MISO channels, an important idea of using SCSI is to precode the transmit symbol along the strongest eigenvector (SE) to the intended user [19]. On the other hand in [5], the optimal precoders for Two-User (TU) multi-antenna BC in extremely low- and high-SNR regime can be computed by closed-form expressions. However, in intermediate-SNR regime, the precoders must be obtained via solving an optimization problem and the optimality is not proved. Nevertheless, the closed-form ergodic group-rate expression of two-user case with arbitrary precoding vectors is derived, which constitutes a significant basis of our simulations. To overcome the complexity of searching optimal precoders, [4] suggests to apply a set of suboptimal beamforming vectors which can be easily found and achieve nearly equivalent sum-rate to [5].

Simultaneously sending K different signals to K users ($K \geq 2$) usually requires the number of antennas to be larger than or equal to the number of users [4], [5], [15]. Otherwise, due to the limitation of spatial degrees-of-freedom (DoF), the communication system has to consume DoF in other domain. However along with the increment of antennas requirement, the transmission scheme can be unfavorably complex and the precoder design is more difficult. Even in a K transmit antennas K users case with SCSI only, communicating with all users at the same time is inefficient because each link suffers severe inter-user interference. It is noticed that the inter-user interference is inevitable because of following reasons: first, transmit correlation matrices are never rank-deficient [20], which means we cannot find precoders that perfectly

eliminate the interference from other users; second, even if we assume the minimum eigenvalue of a user's correlation matrix is small enough such that we may ignore it and we project other users' signals onto the null space, interference still exists due to the uncertainty of CSI. Nevertheless, in the same system model of our study, [21], [22] conduct research based on an approximated expression of signal-to-interference-and-noise ratio (SINR), where [21] lets the BS simultaneously transmit data to all users with a goal to maximize the minimum SINR among the users and [22] focuses on maximizing the weighted sum-rate (WSR). In addition, [23] proposes a novel transmission scheme that can achieve high sum-rate by ingenious but complicated non-linear encoding and decoding processes.

To avoid inter-user interference and system complexity, we stick to one-user or two-user mode while we still attempt to increase the ergodic sum-rate via user scheduling. For instance, in [15] a Round-Robin fashion (RRF) user scheduling is used to deal with MU scenarios. Also in [12] an adaptive grouping algorithm could achieve higher data-rate per user than RRF. Hence we notice in MU scenarios, by proper user scheduling, the throughput of the communication system could be effectively improved. To this topic, there are two related works [24], [25] which choose precoders in [19] and apply Munkres algorithm [26] [27] to handle the user scheduling problem. However, in those two works, the authors laid a constraint that each user could only be allocated to one group. Thus, the scheduling problem actually transforms into a general maximum weight matching problem in graph theory, which can be solved by [28]. In this paper, we explore the possibility to further enhance systematic performance via user scheduling. Against the current literature, the main contributions of this paper are as follows:

- To avoid inter-user interference and to increase average group-rate of the communication system, we firstly propose a novel user scheduling algorithm (T2-MUSA) based on SCSI. In our algorithm, the energy consumed for each user is the same and all users are allowed to be allocated into more than one user group. Compared with existing algorithms, T2-MUSA realizes higher group rates and achieves considerably satisfactory fairness.
- We show that the achievable group rate when the users can be allocated into more than two groups is upper bounded by that of our proposed T2-MUSA. In short, we prove the optimality of our algorithm. We find a lower bound of the ergodic group-rate, when the dominant eigenvalue of each user's transmit correlation matrix is uniformly distributed.
- We propose another user scheduling algorithm (QB-MUSA) which deals with the queue stability problem. In simulations, we show that proposed QB-MUSA can outperform conventional WSR-based user scheduling algorithms, especially when the arrival rate of each user is comparable to others.

The remainder of this paper is arranged as follows. Section II elaborates the system model of MU-MISO BC in addition with precoder design and problem formulation. In Section III,

we propose algorithm TU-MUSA and discuss the optimality and rate approximation. In Section IV, we list the steps of the proposed algorithm QB-MUSA to solve the queue stability problem. Section V shows numerical results and we summarize the paper in Section VI. Some proofs are presented in Appendices.

Notations: We use uppercase and lowercase boldface letters to denote matrices and vectors respectively. Matrices transpose and Hermitian transpose are denoted by $(\cdot)^T$ and $(\cdot)^H$ respectively. The elementwise conjugate of a complex number or matrix is denoted by $(\cdot)^*$. Expectation operator is denoted by $E[\cdot]$. $\mathcal{CN}(0,1)$ and $U(a,b)$ stand for zero-mean unit variance complex Gaussian distribution and uniform distribution within the interval (a,b) correspondingly. $[\mathbf{M}]_{i,j}$ denotes the (i,j) entry of the matrix. $\|\cdot\|$ is the Frobenius norm operation. $\mathbf{u}_{\max}(\cdot)$ and $\mathbf{u}_{\min}(\cdot)$ denote the eigenvector corresponding to the largest and least eigenvalue $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ respectively. $\mathrm{tr}(\cdot)$ and $\mathrm{rank}(\cdot)$ denote the trace and rank of a matrix. $\mathrm{mod}(X,Y)$ returns the modulus after division of X by Y. $\mathbf{I}$ is reserved for the identity matrix while $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{C}$ and $\mathbb{R}$ are reserved for natural number (including zero), integer complex number and real number set respectively.

II. System Model

We consider MU-MISO *Rayleigh fading* BC where the base station is equipped with N_t antennas ($N_t \geqslant 2$) and each user has a single antenna, see Fig. 1. We define $\mathbb{G}$ as a set of all associated users of a BS and without loss of generality, if there are K users, it can be expressed as $\mathbb{G} = \{1,2,...K\}$. IPCSI is only available at receiver side while SCSI is known to both ends. The channel vector between user k and BS is denoted as $\mathbf{h}_k$ and can be written as

$$\mathbf{h}_k = \mathbf{g}_k \mathbf{R}_k^{\frac{1}{2}} \tag{1}$$

where $\mathbf{g}_k \in \mathbb{C}^{1 \times N_t}$ is independently and identically distributed (i.i.d) $\sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. $\mathbf{R}_k$ is the long-term observed transmit correlation matrix (SCSI) of user k that can be eigenmode decomposed as

$$\mathbf{R}_k = E[\mathbf{h}_k^H \mathbf{h}_k] = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^H \tag{2}$$

where $\mathbf{U}_k \in \mathbb{C}^{N_t \times N_t}$ is a unitary matrix whose columns are the eigenvectors of $\mathbf{R}_k$ and $\mathbf{\Lambda}_k = \mathrm{diag}(\lambda_{k1}, \lambda_{k2}, ...\lambda_{kN_t})$ is a diagonal matrix whose diagonal entries are the corresponding eigenvalues normalized to $\mathrm{tr}(\mathbf{\Lambda}_k) = N_t$. The reason why we can make this normalization is due to the fact that $\mathbf{R}_k$ is a positive definite Hermitian matrix i.e. each eigenvalue is a positive number. Particularly, $\mathbf{\Lambda}_k = \mathbf{I}$ means that the link towards user k is spatially uncorrelated while $\mathrm{rank}(\mathbf{\Lambda}_k) = 1$ indicates the link is fully correlated [29].

On the BS side, we precode signals with $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ before transmitting them where two modes could be opted namely, SU mode and TU mode. For SU mode the BS sends one symbol to one user, while for TU mode it simultaneously transmits to two users with each user receiving one symbol. A user group can then be defined as $\pi_i = \{(i_1, i_2, ..., i_{N_i})|i_1, i_2, ..., i_{N_i} \in \mathbb{G}\}$, where N_i is the size of this user group. A user schedule is a set of user groups
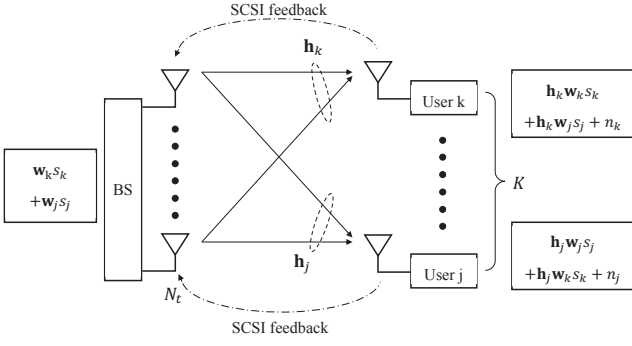
Fig. 1: System model

which the BS will send data to each one of them periodically, $S = \{\pi_1, \pi_2, ...\pi_L\}$ (L is the length). It is noteworthy that S is a multiset which allows duplicated elements.

Throughout this paper, we only consider $N_i = 1, 2$ (SU or TU) and user schedules containing user groups with $N_i > 2$ are beyond the scope of our work. Suppose we have a user group $\pi_i = (k, j)$, the received signal of user k is

$$y_k = \sqrt{\rho}\mathbf{h}_k\mathbf{w}_k s_k + \sqrt{\rho}\mathbf{h}_k\mathbf{w}_j s_j + n_k \quad (3)$$

where $\rho = P/N_i$ and P is the total transmit power constraint, s_k denotes the signals intended for user k with $|s_k|^2 = 1$, $n_k \sim \mathcal{CN}(0, 1)$ stands for normalized complex additive white Gaussian noise. If π_i is in SU mode, we omit the second term of (3) and the associated terms in the sequel. Thus the signal-to-interference-and-noise ratio (SINR) can be found as

$$\text{SINR}_k = \frac{\rho|\mathbf{h}_k\mathbf{w}_k|^2}{\rho|\mathbf{h}_k\mathbf{w}_j|^2 + 1}. \quad (4)$$

Consequently, we have the group-rate[1] of user group π_i

$$R_{\pi_i} = \sum_{k \in \pi_i} R_{k|\pi_i} = \sum_{k \in \pi_i} E\left[\log\left(1 + \text{SINR}_k\right)\right] \quad (5)$$

and the average group-rate of a user schedule

$$\bar{R}_S = \frac{1}{N_i} \sum_{\pi_i \in S} R_{\pi_i}. \quad (6)$$

where $R_{k|\pi_i}$ is the achievable rate of user k in π_i. According to [5], we have the closed-form expression of $R_{k|\pi_i}$ for both

---

[1]To be consistent with [5], we use all logarithms to base e. Hence in all calculations, the rate unit is nats/s/Hz whereas in graphs we multiply results by $\log_2 e$ to convert the unit back into bits/s/Hz.

SU mode and TU mode. Specifically if π_i is in SU mode, it is given by

$$R_{k|k} = h(\rho A_k) \quad (7)$$

where

$$A_k = \mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k \quad (8)$$

$$h(x) \triangleq \exp(\frac{1}{x})E_1(\frac{1}{x}), \quad x \in (0, \infty) \quad (9)$$

with $E_1(\cdot)$ being the Exponential integral function, namely $E_1 = \int_x^\infty \frac{e^{-t}}{t}dt$. On the other hand, suppose we have two users (k and j) in π_i, the closed-form expression of user k's achievable rate can be written as

$$R_{k|(k,j)} = \frac{\mathbf{\Lambda}_{k,1}h(\rho\mathbf{\Lambda}_{k,1}) - \mathbf{\Lambda}_{k,2}h(\rho\mathbf{\Lambda}_{k,2})}{\mathbf{\Lambda}_{k,1} - \mathbf{\Lambda}_{k,2}} - h(\rho B_k) \quad (10)$$

where

$$\mathbf{\Lambda}_{k,1} = \frac{1}{2}\left(A_k + B_k + \sqrt{(A_k - B_k)^2 + 4C_k^2}\right) \quad (11)$$

$$\mathbf{\Lambda}_{k,2} = \frac{1}{2}\left(A_k + B_k - \sqrt{(A_k - B_k)^2 + 4C_k^2}\right) \quad (12)$$

with A_k the same as (8) and $B_k = \mathbf{w}_k^H \mathbf{R}_j \mathbf{w}_k$, $C_k = |\mathbf{w}_k^H \mathbf{R}_j \mathbf{w}_j|$. What's more, we also have to define

$$R_{k|\pi_i} = 0, \quad k \notin \pi_i. \quad (13)$$

A. Statistical Precoders

We now analyze the precoder design for both SU and TU modes. When choosing precoders, we tend to maximize the group-rate of the corresponding user group and simplify the implementation complexity. In SU mode, we have the SNR of user k as

$$\text{SNR}_k = \rho|\mathbf{h}_k\mathbf{w}_k|^2 = \rho\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k, \quad (14)$$

therefore, the optimal precoder of a single user (OPSU) should be along the strongest eigenvector of the transmit correlation matrix [19] i.e.

$$\mathbf{w}_k = \mathbf{u}_{\max}\left(\mathbf{R}_k\right). \quad (15)$$

This precoder structure is applied to every user throughout the works of [24], [25] for both SU and TU mode, although it is not actually designed for TU groups. On the other hand, [5] provides the optimal precoders for TU mode (OPTU) and these precoders have different expressions in different SNR regime (see TABLE I). To simplify implementation complexity [4] proposes a set of sub-optimal precoders for

TABLE I: TU Mode Precoders Comparison

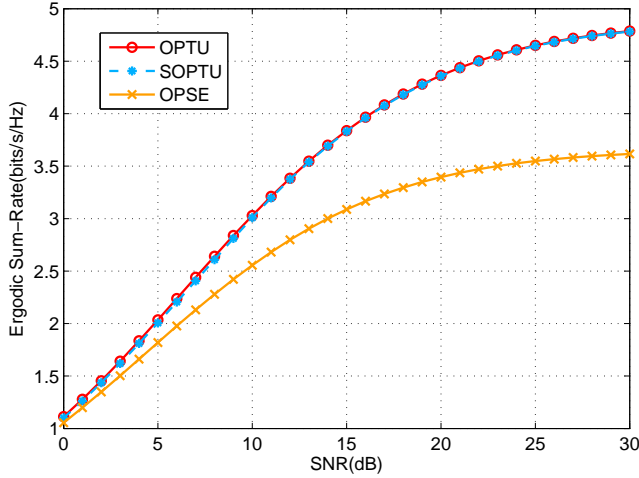|	OPTU [5]	SOPTU [4]	OPSU [19], [24], [25]
Low SNR	$\mathbf{w}_k = \mathbf{u}_{\max}\left(\mathbf{R}_k\right)$ $\mathbf{w}_j = \mathbf{u}_{\max}\left(\mathbf{R}_j\right)$		
Intermediate SNR	$\mathbf{w}_k = \mathbf{u}_{\max}\left((\alpha(\rho)\mathbf{R}_j + \mathbf{I})^{-1}\mathbf{R}_k\right)$ $\mathbf{w}_j = \mathbf{u}_{\max}\left((\beta(\rho)\mathbf{R}_k + \mathbf{I})^{-1}\mathbf{R}_j\right)$	$\mathbf{w}_k = \mathbf{u}_{\max}\left((\rho\mathbf{R}_j + \mathbf{I})^{-1}\mathbf{R}_k\right)$ $\mathbf{w}_j = \mathbf{u}_{\max}\left((\rho\mathbf{R}_k + \mathbf{I})^{-1}\mathbf{R}_j\right)$	$\mathbf{w}_k = \mathbf{u}_{\max}\left(\mathbf{R}_k\right)$ $\mathbf{w}_j = \mathbf{u}_{\max}\left(\mathbf{R}_j\right)$
High SNR	$\mathbf{w}_k = \mathbf{u}_{\max}\left(\mathbf{R}_j^{-1}\mathbf{R}_k\right)$ $\mathbf{w}_j = \mathbf{u}_{\max}\left(\mathbf{R}_k^{-1}\mathbf{R}_j\right)$		

Fig. 2: Sum-rate performance comparison with different TU mode precoders. Results are averaged over 500 realizations of $N_t = 2$, $|t_k| \in U(0,1)$ and $\phi_k \in U(0,2\pi)$.

TU mode (SOPTU) which can be found by a simple closed-form solution. As ρ approaches 0 or infinity, SOPTU precoders converge to the OPTU precoders and in the intermediate SNR regime, SOPTU can be seen as a special case where $\alpha(\rho) = \rho$ and $\beta(\rho) = \rho$. However, OPSU precoders keep consistent with the structures of low-SNR OPTU precoders which are irrelevant with the individual transmit power ρ. In Fig. 2 we compare the performance of these three precoders at different SNR levels where the BS is equipped with 2 transmit antennas. The simulation result is averaged over 500 random TU groups with each user generated by the single parameter exponential correlation model [30] namely,

$$\mathbf{R}_k = \begin{bmatrix} 1 & t_k & \cdots & t_k^{M-1} \\ t_k^H & 1 & \cdots & t_k^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ (t_k^H)^{M-1} & (t_k^H)^{M-2} & \cdots & 1 \end{bmatrix} \quad (16)$$

where the correlation coeficient $t_k = |t_k| e^{j\phi_k}$ is randomly generated by $|t_k| \in U(0,1)$ and $\phi_k \in U(0,2\pi)$.

Remark 1: To the best of our knowledge, the closed-form statistical optimal precoders for TU mode have not been found yet. SOPTU precoders demonstrate a comparable performance to OPTU, although in some rare cases the group-rate difference between these two might be distinct. OPTU and SOPTU outperform OPSU in all SNR regimes. Therefore, in this paper we choose the SOPTU structure as the TU mode precoders.

B. Problem Formulation

Maximizing the average group-rate leads to one solution that the user schedule only consists of the optimal user group. However, this schedule is impractical since we have to consider sending data to all associated users rather than some of them. To this end, we put some constraints on the user schedule and first of all, it should be a complete schedule whose definition is given below.

Definition 1: A complete user schedule includes all users in $\mathbb{G}$.

Second, the user scheduling should be 'fair' to all users in a certain sense and there are two factors we mainly take into consideration in terms of fairness issue i.e. frequency and energy. In contrast with the conventional meaning in communication engineering, 'frequency' here means the number of appearances of an individual user in a schedule. Basically speaking, a fair algorithm of a certain resource lets all users equally share this resource. We define $\mathbb{S}_{FF}$ as the space of frequency-fair user schedules and $\mathbb{S}_{EF}$ as the space of energy-fair schedules. [24] and [25] give two frequency-fair user schedules algorithms (FFUSA) with almost equivalent performance. In these two FFUSA the frequency of each user is strictly limited to 1 and this constraint actually has a negative effect on the performance when user number is an odd number. We will show and discuss this effect in the Numerical Results part. Additionally, it can be easily seen that the RRF schedule belongs to both frequency-fair and energy-fair schedules, $S_{RRF} \in \mathbb{S}_{EF} \cap \mathbb{S}_{FF}$. Hence the first problem we try to solve in this paper is to find a user scheduling algorithm which leads to an energy-fair or frequency-fair or even better, both fair schedule that maximizes the average group sum-rate in (6) namely

$$\text{P1}: \quad \arg\max_{S \in \mathbb{S}_{EF} \cup \mathbb{S}_{FF}} \bar{R}_S. \quad (17)$$

In the above problem, we do not involve other factors that might give priorities to some users, which means we conceive that all users are equally important and require presumably equivalent quality of services. Hence, we consider the fairness metric in terms of the data-rate of individual users. In details, we compute the achievable rate of user k in a user schedule as

$$R_{k|S} = \sum_{\pi_i \in S} R_{k|\pi_i}. \quad (18)$$

And then we make the Jain's fairness index [31] as our fairness metric, namely

$$\mathcal{F} = \frac{\left(\sum_{k=1}^{K} R_{k|S}\right)^2}{K \sum_{k=1}^{K} R_{k|S}^2}. \quad (19)$$

This metric ranges from $1/n$ to 1, where $1/n$ stands for the worst case that only one user occupies the whole channel and 1 stands for the best case that every user has the same data-rate.

From another point of view, the meaning of user scheduling is to stabilize queue lengths. Specifically, if we denote $A_k(t)$ as the arrival process, $Q_k(t)$ as the queue length and $R_k(t)$ as the data-rate to user k in time slot t, we have the queue length in the next time slot as

$$Q_k(t+1) = Q_k(t) - R_k(t) + A_k(t), \quad (20)$$

where $R_k(t)$ cannot be larger than $Q_k(t)$.

To keep user queues stable is actually to design a feasible user scheduling algorithm $\mathcal{A}$ that solves the minimization problem

$$\text{P2}: \quad \arg\min_{\mathcal{A}} \sum_{k=1}^{K} Q_k(t), \quad \forall t, \quad (21)$$

so that the data stored on the BS side will not exceed the buffer size. To achieve this goal, both high data-rates and a balance over user queues are expected. Conventionally, this problem can be solved by weighted sum-rate (WSR) maximization [22], where queue lengths may be directly used as the weight coefficients for user data-rates, so that users with a long queue length have advantage in user scheduling. We notice that this kind of scheduling algorithm is *greedy*, because they focus on maximizing momentary data-rate and queue reduction. However the problem we concerned is a long term optimization problem. The fairness in terms of frequency or energy in this problem therefore is not important, as the balance of user queues stand for another kind of fairness.

III. ENERGY-FAIR USER SCHEDULING ALGORITHM

In this section we focus on the problem P1 and we consider to release the frequency of each user to any positive integer, in other words, a user can be in more than one group. For instance, the following user schedule is a frequency-fair schedule with each user's frequency equal to 2.

$$S = \{(1), (3), (1,2), (2,4), (4,3)\}. \tag{22}$$

To gain the maximum throughput, we assume that the BS inherently work with full power no matter in SU mode or TU mode. In addition, we make the following definition.

Definition 2: The shortest complete user schedule that satisfies

$$S^\star = \underset{S \in \mathbb{S}_{EF} \cap \mathbb{S}_{FF}}{\arg\max} \bar{R}_S \tag{23}$$

is defined as the optimal fair schedule.

We emphasize that there are infinite number of schedules satisfying (23) e.g. a schedule that consists of duplicated copies of $S^\star$ namely $S = \{S^\star, S^\star, ..., S^\star\}$. Therefore, we define the optimal fair schedule as the shortest one and then Theorem 1 is found true.

Theorem 1: If $S^\star$ is the optimal fair schedule defined in Definition 2, $S^\star$ must be composed of either fully SU groups or fully TU groups.

Proof: See Appendix A.

Since there is only one shortest permutation in terms of complete user schedule with fully SU groups, the optimal fair schedule in this case can be written as

$$S_1^\star = \{(1), (2), ..., (K)\}. \tag{24}$$

Meanwhile, there are many permutations in terms of complete user schedule with fully TU groups. Nevertheless Munkres algorithm is able to efficiently solve this problem within $\mathcal{O}(K^3)$ time so that we have a user group set $S_2^\star$. Therefore (23) becomes

$$S^\star = \underset{S = S_1^\star \text{ or } S_2^\star}{\arg\max} \bar{R}_S. \tag{25}$$

As expressed in (25), the optimal fair schedule can be easily found. Normally, $S_2^\star$ has higher sum-rate than $S_1^\star$ has in low SNR regime while $S_1^\star$ wins in high SNR regime, and this will be discussed in the sequel. However, if higher average group-rate beyond $\bar{R}_{S^\star}$ is expected, we must lose some fairness either in terms of energy or frequency, for example, the mentioned [24] and [25] only focus on frequency fairness.

A. Tier-2 Munkres User Scheduling Algorithm

In our work, we mainly focus on the fairness in energy domain which means each user consumes the same amount of energy while does not necessarily has the same frequency in the schedule. We define a tier-N schedule by that the maximum energy consumed by a user in the schedule is equal to $\frac{NP}{2}$ e.g. the schedule in (22) is a tier-3 schedule. We now propose tier-2 Munkres user scheduling algorithm(T2-MUSA) which is elaborated in Algorithm 1.

Algorithm 1 T2-MUSA

1: Initialize group-rate distribution matrix $\mathbf{D}$ as a $K \times K$ zero matrix i.e. $\mathbf{D} = \mathbf{0}$.
2: **for** $i = 1 : K$ **do**
3:     **for** $j = 1 : K$ **do**
4:         **if** $i = j$ **then**
5:             make the $(i,i)^{th}$ entry equal to the achievable rate of user i in SU mode, $[\mathbf{D}]_{i,i} = R_i$.
6:         **else**
7:             make the $(i,j)^{th}$ entry equal to the group-rate of TU group (i,j), $[\mathbf{D}]_{i,j} = R_{(i,j)}$.
8:         **end if**
9:     **end for**
10: **end for**
11: Take the opposite matrix of the group-rate distribution matrix, $\mathbf{D} = -\mathbf{D}$.
12: Apply Munkres algorithm to the opposite matrix and find the optimal tier-2 energy-fair schedule $S_{EFT2}^\star$.

We show a group-rate distribution matrix with 4 users in (26), and each entry of $\mathbf{D}$ represents the group-rate in either SU or TU mode while energy consumed for every group is the same. Moreover, the definition of the optimal energy-fair schedule is the same as Definition 2 but we reduce the searching domain to $S \in \mathbb{S}_{EF}$ in (23).

$$\mathbf{D} = \begin{bmatrix} R_1 & R_{(1,2)} & R_{(1,3)} & R_{(1,4)} \\ R_{(2,1)} & R_2 & R_{(2,3)} & R_{(2,4)} \\ R_{(3,1)} & R_{(3,2)} & R_3 & R_{(3,4)} \\ R_{(4,1)} & R_{(4,2)} & R_{(4,3)} & R_4 \end{bmatrix}. \tag{26}$$

It is noteworthy that Munkres algorithm is designed for finding a set of K independent entries (Definition 3) of a $K \times K$ matrix such that the sum of these entries is the smallest among all permutations. Therefore, we convert each entry of $\mathbf{D}$ into its opposite number before using Munkres algorithm to find the permutation with maximum sum-rate.

Definition 3: A set of entries of a matrix are independent if none of them occupies the same row or column with others [27].

Specifically, T2-MUSA gives a $K \times K$ assignment matrix $\mathbf{A}$ containing independent '1's that indicate the assigning status of corresponding user groups i.e.

$$[\mathbf{A}]_{ij} = \begin{cases} 0 & \text{user group } (i,j) \text{ is not assigned}, \\ 1 & \text{user group } (i,j) \text{ is assigned}. \end{cases} \tag{27}$$

Remark 2: For a given assignment matrix $\mathbf{A}$, suppose n is the maximum number of independent '1's of $\mathbf{A}$, then we can

find n lines (rows or columns or both) which contain all the '1' elements [26]. (This theorem constitutes one of the important bases of Munkres algorithm.) Besides, we emphasize that in K users cases Munkres algorithm can always find an assignment matrix $\mathbf{A}$ with K independent '1' elements as long as each entry of the group-rate distribution matrix is a real number, which is to say the resulting matrix $\mathbf{A}$ always corresponds to a complete schedule since there must be a '1' element in each row (or column). Actually to a wide sense, any set of K independent '1's forms a complete schedule but $\mathbf{A}$ is the optimal one.

The algorithm does not exclude some frequency-fair schedules e.g. $S_1^\star$ or $S_2^\star$ can be found as long as it is the optimal permutation. Particularly, $S_1^\star$ corresponds to the case $\mathbf{A} = \mathbf{I}$ and $S_2^\star$ satisfies $\mathrm{diag}\{\mathbf{A}\} = \mathbf{0}$. Any assignment matrix except these two special cases stands for a schedule that only lies in energy-fair rather than frequency-fair space.

B. Optimality and Rate Analysis

In this section, we prove the optimality of T2-MUSA in the energy-fair schedule space. By optimality, it means the schedule found by our proposed algorithm achieves highest average group-rate and is with shortest length. Besides, we also derive a lower bound of the average group-rate in a special case.

Lemma 1: For a given group-rate distribution matrix and an arbitrary tier-N schedule, if $M = mN, m \in \mathbb{Z}^+$, we can always find a tier-M schedule such that

$$\bar{R}_{S_{TM}} \geq \bar{R}_{S_{TN}}. \tag{28}$$

Proof: Since at least we can have a tier-M schedule which is composed of m duplicated copies of the tier-N schedule i.e.

$$S_{TM} = \{\underbrace{S_{TN}, S_{TN}, \cdots, S_{TN}}_{m}\}. \tag{29}$$

Hence we complete the proof.

In Lemma 1, we do not specify the tier-N schedule as neither an energy- or frequency-fair schedule. So it works for all schedules even for incomplete user schedules. By this Lemma we can write

$$\bar{R}_{S_{EFT2}^\star} \leq \bar{R}_{S_{EFT2m}^\star}, \quad m = 2, 3, 4 \cdots. \tag{30}$$

where $S_{EFT2m}^\star$ could be efficiently found by applying Munkres algorithm to the $mK \times mK$ extended group-rate distribution matrix

$$\mathbf{D}_m = \underbrace{\begin{bmatrix} \mathbf{D} & \mathbf{D} & \cdots & \mathbf{D} \\ \mathbf{D} & \mathbf{D} & \cdots & \mathbf{D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D} & \mathbf{D} & \cdots & \mathbf{D} \end{bmatrix}}_{m}. \tag{31}$$

Specifically, we can find an $mK \times mK$ assignment matrix $\mathbf{A}_m$ with mK independent '1' elements using Munkres algorithm. For each '1' element in $\mathbf{A}_m$, take the one located at the $(i, j)^{th}$ entry as an example, we make the coordinate transfer by

$$\begin{cases} i' = \mathrm{mod}(i - 1, K) + 1, \\ j' = \mathrm{mod}(j - 1, K) + 1. \end{cases} \tag{32}$$

Then we add the group (i', j') into the schedule and in this way we build up the optimal tier-$2m$ energy-fair schedule $S_{EFT2m}^\star$.

Lemma 2: Suppose $\mathbf{B}$ is the resulting matrix by applying Mapping algorithm in Separation Step 1 to the assignment matrix $\mathbf{A}_m$, there must exist K independent '1's in $\mathbf{B}$.

Separation Step 1 Mapping algorithm

1: Initialize $\mathbf{B}$ as a $K \times K$ zero matrix, $\mathbf{B}=\mathbf{0}$.
2: **for** $i = 1 : mK$ **do**
3:     **for** $j = 1 : mK$ **do**
4:         **if** $[\mathbf{A}_m]_{i,j} = 1$ **then**
5:             Make the coordinate transfer by (32) and obtain updated coordinate i' and j'.
6:             Change the corresponding bit in $\mathbf{B}$ as $[\mathbf{B}]_{i',j'} = 1$.
7:         **end if**
8:     **end for**
9: **end for**

Proof: See Appendix B for detailed proof.

Separation Step 2: It is notable that some '1' elements of $\mathbf{A}_m$ may refer to the same entry in $\mathbf{B}$ by the mapping algorithm e.g. the '1' elements positioned at $(1, 1)^{th}$ and $(K + 1, K+1)^{th}$ entry of $\mathbf{A}_m$ are both mapped to the $(1, 1)^{th}$ entry of $\mathbf{B}$. In other words, a '1' element in $\mathbf{B}$ may be associated with many '1's of $\mathbf{A}_m$. We now denote a matrix $\mathbf{C}_1$ containing K independent '1's in $\mathbf{B}$ and for each '1' element in $\mathbf{C}_1$, we find its associated '1' element in $\mathbf{A}_m$ (if there are more than one associated '1's, we randomly choose one of them). Then, we remove the row and column where the associated '1' element lies from $\mathbf{A}_m$. By doing so, we remove K rows and K columns from $\mathbf{A}_m$ and thereby form a new $(m-1)K \times (m-1)K$ square matrix denoted by $\mathbf{A}_{m-1}$. It is plain to see that removing rows or columns from a assignment matrix does not destroy the independence of '1' elements, therefore, we have $(m - 1)K$ independent '1's in $\mathbf{A}_{m-1}$ which means it forms a tier-$2(m - 1)$ complete schedule $S_{T2(m-1)}$. Together with the tier-2 complete schedule formed S_{T2}^1 by $\mathbf{C}_1$, we actually separate the optimal energy-fair schedule $S_{EFT2m}^\star$ into two complete schedules i.e.

$$S_{EFT2m}^\star = \{S_{EFT2(m-1)}, S_{EFT2}^1\}. \tag{33}$$

Corolarry 1: The optimal tier-$2m$ energy-fair schedule $S_{EFT2m}^\star$ can be separated into m tier-2 complete schedules, in mathematics we have

$$S_{EFT2m}^\star = \{S_{EFT2}^1, S_{EFT2}^2, \cdots, S_{EFT2}^m\}. \tag{34}$$

Proof: With the help of Remark 2 and Lemma 2, it can be easily envisioned that if we repeat the procedure of Separation Step 1 and 2, (34) could be eventually achieved, which completes the proof.

Here we give a simple example of what is done in the Separation process. Suppose we have 3 users and we have a tier-4 assignment matrix in (35) whose size is 6×6 and

corresponding arrangement (using coordinate transfer (32)) is given by (36), i.e.

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{35}$$

$$S_{EFT4} = \{(1,2),(2,3),(3),(1,2),(2,1),(3,1)\}. \tag{36}$$

As we can see, the energy expended by each user is $2P$ which is consistent with a tier-4 energy-fair arrangement. Now we apply Separation Step 1 to map $\mathbf{A}_2$ into $\mathbf{B}$ and get

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}. \tag{37}$$

Thus we can find 3 independent '1' elements in $\mathbf{B}$ and put them into $\mathbf{C}_1$. Then, we remove the corresponding '1' elements in $\mathbf{A}_2$ and their columns and rows and get $\mathbf{A}_1$,

$$\mathbf{C}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{38}$$

Now, $\mathbf{C}_1$ and $\mathbf{A}_1$ form two tier-2 energy-fair arrangements and we successfully separate S_{EFT4} by

$$S_{EFT4} = \underbrace{\{(1,2),(2,3),(3,1)\}}_{\mathbf{C}_1} + \underbrace{\{(1,2),(2,1),(3)\}}_{\mathbf{A}_1}. \tag{39}$$

With Lemma 1 and Corolarry 1, we are ready to prove the optimality of T2-MUSA.

Theorem 2: The optimal tier-2 energy-fair schedule is the optimal energy-fair schedule among all tier-N schedules where $N \geq 2$.

Proof: Refer to Appendix C for proof.

Although $S_{EFT2}^\star$ is proved to be the optimal schedule in energy-fair schedule space, its optimality in the joint space of energy-fair and frequency-fair schedule spaces cannot be guaranteed. However, in ergodic experiments, T2-MUSA demonstrates its superiority over frequency-fair algorithms and this will be confirmed in our later simulations.

In practical wireless transmission, the transmit correlation matrix is never rank-deficient which means the eigenvalues should always be larger than zero. According to the SINR found in (4), we know that the inter-user interference inherently exists in TU mode transmissions. In other words, we cannot find a precoder that could manage to perfectly eliminate this interference since at the very least it should be no less than $\rho\lambda_{min}$, where λ_{min} is the minimum eigenvalue of the other user in the TU group. For this reason, the group-rate finally saturates to a certain level as SNR increases and this phenomenon is well discussed in [4]. Meanwhile for a SU group, the transmitted signals never suffer interference and hence the achievable rate goes up along with the transmit power. We can easily find the DoF of SU mode rate by

$$\mathrm{DoF}_{su} = \lim_{P \to \infty} \frac{\log(1 + P\lambda_{max})}{\log P} = 1. \tag{40}$$

Hence in extremely high SNR regime, due to that TU mode transmission severely underperforms SU mode, both energy-fair and frequency-fair schedules converge to all-SU-group schedules i.e. $S_1^\star$ (24). As a result of this situation, we can deduce the lower bound of the average group-rate of our proposed algorithm and it is also the lower bound of frequency-fair algorithms.

Proposition 1: If the maximum eigenvalues of the statistical transmit correlation matrices are uniformly i.i.d within the interval (a, b) such that $1 \leq a \leq b \leq N_t$, then the average group-rate of $S_{EFT2}^\star$ is lower bounded by

$$\bar{R}_{S_{EFT2}^\star} \geq \frac{1}{P}\left(g\left(bP\right) - g\left(aP\right)\right). \tag{41}$$

where $g(x)$ is defined in (42) with γ being the Euler-mascheroni constant.

$$g\left(x\right) = \frac{\log^2 x}{2} + (x + 1 - \gamma)\log x - (\gamma + 1)x \\ + \frac{3}{4x} + \frac{1}{x^2}, \qquad x \in (0, \infty). \tag{42}$$

Proof: See Appendix D for detailed proof.

As we discussed before, SU mode transmissions only have the rate superiority in extremely high SNR regime. Therefore, the lower bound (42) is loose in low SNR regime and turns to tight as SNR rises.

IV. QoS-Based User Scheduling Algorithm

In this section, we propose a user scheduling algorithm inspired by T2-MUSA and aimed to solve problem P2. we suppose the initial queue lengths (when $t = 0$) of all users are written as $\mathbf{i} \in \mathbb{R}_+^K$ and Let $\mathbf{r}(t) = [R_1(t), R_2(t), \ldots, R_K(t)]^T$ denote the users' data-rate set. Besides, $\mathcal{R}(t)$ denotes the region of rates which depicts the space that $\mathbf{r}(t)$ can take values from and actually

$$\mathcal{R}(t) = \\ \bigcup_{\forall \pi_i} \left\{ \mathbf{a} \in \mathbb{R}^K \Big| [\mathbf{a}]_k = \min\left\{ R_{k|\pi_i}, Q_k(t) \right\}, k = 1, 2, ..., K \right\} \tag{43}$$

gives the mathematical definition. Since we only consider TU or SU grouping, $\mathbf{r}(t)$ has at most two non-zero entries in a particular time slot. So far, we can make the following equation,

$$\sum_{k=1}^{K} Q_k(T) = \|\mathbf{i}\| + \sum_{t=1}^{T}\sum_{k=1}^{K} A_k(t) - \sum_{t=1}^{T} \|\mathbf{r}(t)\| \tag{44}$$

(recall that Q_k and A_k denote the ongoing queue length and arrival process). Because the initial queue lengths and arrival process are independent to our user scheduling algorithm, the problem of minimizing the amount of queueing data can be equivalently converted into maximizing the throughput over time i.e.

$$\arg\max_{\mathbf{r}(t) \in \mathcal{R}(t)} \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{r}(t)\|. \tag{45}$$

To solve the above problem, we borrow the idea of using queue lengths to represent importance (weights) of users and

propose a QoS-Based Munkres user scheduling algorithm (QB-MUSA) as in Algorithm 2.

Algorithm 2 QB-MUSA

1: Initialize group-rate distribution matrix $\mathbf{D}$ as a $K \times K$ zero matrix i.e. $\mathbf{D} = \mathbf{0}$.
2: **for** $i = 1 : K$ **do**
3:    **for** $j = 1 : K$ **do**
4:       **if** $i = j$ **then**
5:          make the $(i,i)^{th}$ entry equal to the achievable rate of user i in SU mode, $[\mathbf{D}]_{i,i} = Q_i(t)R_i$.
6:       **else**
7:          make the $(i,j)^{th}$ entry equal to the group-rate of TU group (i,j), $[\mathbf{D}]_{i,j} = Q_i(t)R_{i|(i,j)} + Q_j(t)R_{j|(i,j)}$.
8:       **end if**
9:    **end for**
10: **end for**
11: Take the opposite matrix of the group-rate distribution matrix, $\mathbf{D} = -\mathbf{D}$.
12: Apply Munkres algorithm to $\mathbf{D}$ and find K user groups and K associated independent entries $\mathbf{k} \in \mathbb{R}_-^K$.
13: Find the minimum entry of $\mathbf{k}$ and schedule the corresponding user group for next time slot.

In order to make performance comparison, a greedy user scheduling algorithm (GUSA) is designed. However, we find out that the user scheduling algorithm which solves the WSR problem namely $\max \sum_{k=1}^{K} W_k R_k$ in our defined system model remains unknown. Although [22] attempted this problem, the SNR used in the paper is an approximation. We then make the GUSA always schedule the user group with minimum value in $\mathbf{D}$. As we stated in the introduction part, the inter-user interference is inevitable in SCSI channels due to the nonsingularity of transmit correlation matrices and the uncertainty of channel instance. As more users join the group, the inter-user interference would increase rapidly. So, we consider the GUSA would act in a very similar manner as the algorithm in [22].

The intuition behind QB-MUSA is that Munkres algorithm has long-term consideration for user schedules. However GUSA is kind of short-sighted, only focusing on momentary sum-rate maximization while ignoring the subsequent effects. Since the energy consumed in each entry is the same, Munkres algorithm always produce energy-fair schedules, whereas 'energy-fair' is not a constraint in problem P2. Hence we overcome the drawback by only choosing the user group with maximum weighted group-rate (step 13) in the schedule instead of accepting the whole resulting schedule. However, the superiority of QB-MUSA over GUSA is not guaranteed.

V. NUMERICAL RESUTLS

In this section we provide numerical results to show the performances of our proposed algorithms. The user transmit correlation matrices are in the form of the single parameter exponential model (16) and we apply the SOPTU structure as our precoders as mentioned in Table I. For comparison, we
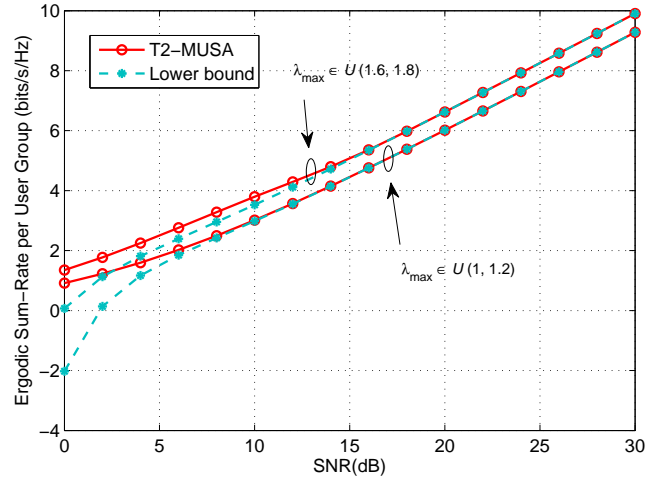

Fig. 3: Comparison between T2-MUSA group-rate and Lower bound in Proposition 1 with 500 realizations of $N_t = 2$, $K = 10$, $|t_k| \in U(1, 1.2)$ or $U(1.6, 1.8)$ and $\phi_k \in U(0, 2\pi)$.

choose the FFUSA in [24] where OPSU precoders are applied. However in this paper we only focus on the user scheduling algorithms. To make fair comparison, we change the precoders of FFUSA to SOPTU structure as well, so that we can get rid of the influence by precoders. All group-rates are obtained by the closed-form expressions given by [5]. Besides, a large family of transmit correlation matrix models are tested in our study to support the superiority of T2-MUSA. In Figures, RRF denotes Round-Robin Fashion user schedules that include every possible TU groups but no SU groups and they can be regarded as normal schedules with no algorithm applied. In addition we also compare our proposed algorithm with the simultaneous transmission scheme in [21] with individual SINR constraints (ISC). We set the individual target SINRs of ISC to 1, so the algorithm tends to balance each user's SINR to the same maximum level.

Since the ISC scheme finds precoders via a optimization process, the complexity is non-polynomial. The difference of complexity between T2-MUSA and FFUSA lies in the user scheduling algorithm, because generating the group-rate distribution matrix (26) should be prepared for both algorithms. The computational complexity for Munkres algorithm is known by $\mathcal{O}(K^3)$. While, for the general maximum weight matching problem, the fastest algorithm yet also needs a $\mathcal{O}(K^3)$ time to solve the problem [28].

A. Lower Bound of Group-Rate

We first present the comparison between the lower bound in Proposition 1 and the simulation results. Since the proposition only works when the maximum eigenvalue is uniformly distributed, we consider the number of transmit antennas $N_t = 2$, the number of users $K = 10$ and the correlation coefficient $|t_k| \in U(a - 1, b - 1)$ and $\phi_k \in U(0, 2\pi)$ where a and b satisfy the condition in the proposition. Thus after some trivial mathematical manipulations, it can be found that the maximum eigenvalue follows the distribution $\lambda_{max} \in U(a, b)$. As shown in Fig. 3, we test two intervals i.e. $\lambda_{max} \in U(1, 1.2)$ and
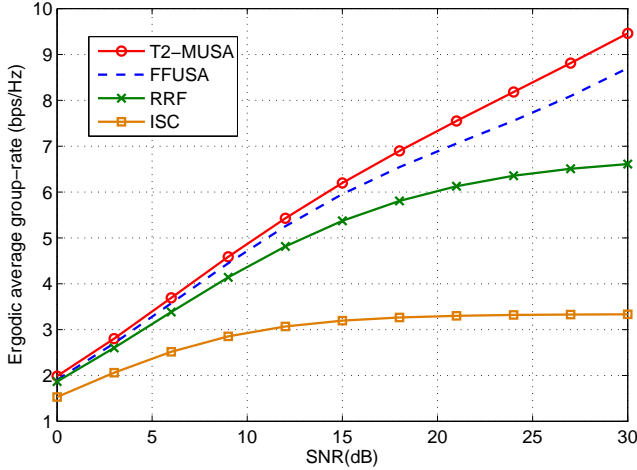
Fig. 4: Comparison of group-rates versus SNR between T2-MUSA and conventional algorithms. Results are averaged over 500 realizations of $N_t = 8$, $K = 15$, $|t_k| \in U(0,1)$ and $\phi_k \in U(0,2\pi)$.
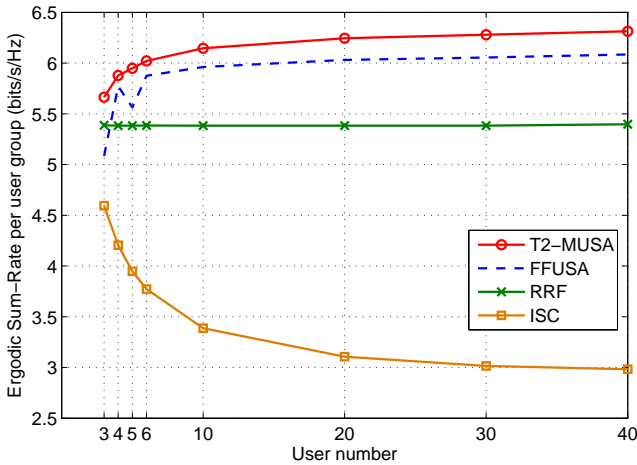

Fig. 5: Comparison of group-rates versus multi-users with 1000 realizations of $N_t = 8$ and power fixed at $P = 15$ dB, $|t_k| \in U(0,1)$ and $\phi_k \in U(0,2\pi)$.

$\lambda_{max} \in U(1.6, 1.8)$ and it shows in these two cases the lower bounds are tight in high SNR regime. In addition, this figure also shows the validity of our simulations.

B. Average Group-Rate Comparison in Different SNRs

The performance of user scheduling algorithms varies from different numbers of transmit antennas, numbers of users and transmit correlation matrix models. Fig. 4 shows the simulation results of transmit antennas $N_t = 8$ and user number $K = 15$. The group-rate curves are averaged over 500 user-set realizations and in each realization, we randomly generate 15 users. The ergodic group-rate of the RRF user schedule actually represents the expected group-rate of an arbitrary TU group. As can be seen from the figure, T2-MUSA and FFUSA effectively improve the ergodic group-rate compared with RRF. The group-rate of T2-MUSA is higher than FFUSA especially in high SNR regime. When SNR reaches around 25 dB, there is a 3 dB gap in terms of transmitting power between T2-MUSA and FFUSA. However
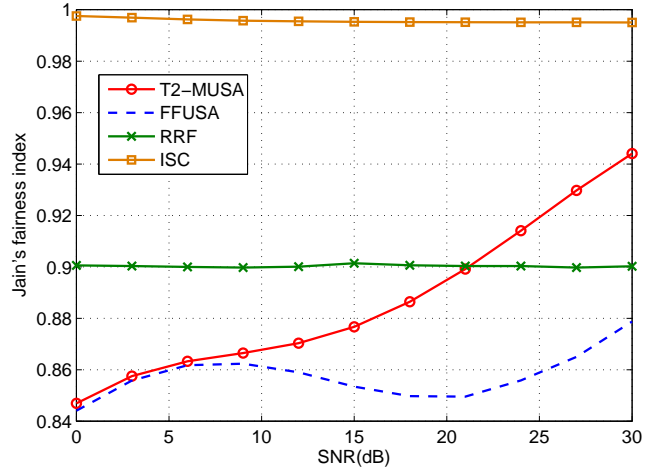

Fig. 6: Fairness comparison over different SNRs with 500 realizations of $N_t = 8$, $K = 15$, $|t_k| \in U(0,1)$ and $\phi_k \in U(0,2\pi)$.

ISC demonstrates the worst performance among all algorithms, which actually shows an example of that if the BS tends to simultaneously send data to more users, the overall data-rates would degrade instead.

C. Average Group-Rate Comparison in Different Numbers of Users

In Fig. 5, we still compare the ergodic group-rates of T2-MUSA with other algorithms but in terms of different numbers of users. The BS is equipped with 8 antennas and the transmit power is fixed at $P = 15$ dB. As RRF scheduling implies expected group-rate of TU groups, it does not benefit from multiuser diversity. On the contrary, ISC suffers from the increment of users as the inter-user interference increases along with rising of the number of users. T2-MUSA outperforms other algorithms, however as the number of users rises, all algorithms tend to saturate at some certain group-rates. An interesting phenomenon is observed from the curve of FFUSA, that is, the performance degrades when the number of users is odd. It arises from the frequency constraint upon the algorithm i.e. each user can only be allocated to one user group, which means, they must choose one user to be in SU group when the total number of users is odd. Meanwhile, in low SNR regime where usually TU mode transmissions have the advantage in group-rate, if a user schedule must include a SU group, the data-rate performance has to thus degrade. However T2-MUSA is immune to this effect, because it can arrange every user into TU groups if it is preferable.

D. Fairness Comparison

In Fig. 6 we focus on the fairness issue of our proposed algorithm and existing algorithms. Due to that FFUSA, T2-MUSA and even ISC respectively achieve their fairness in different aspects, it is hard to tell which one is fairer. So we ignore the differences among these algorithms in energy consumption or user frequencies or individual SINR and we consider the fairness in the view of achievable rates of individual users. As mentioned before, we adopt Jain's fairness
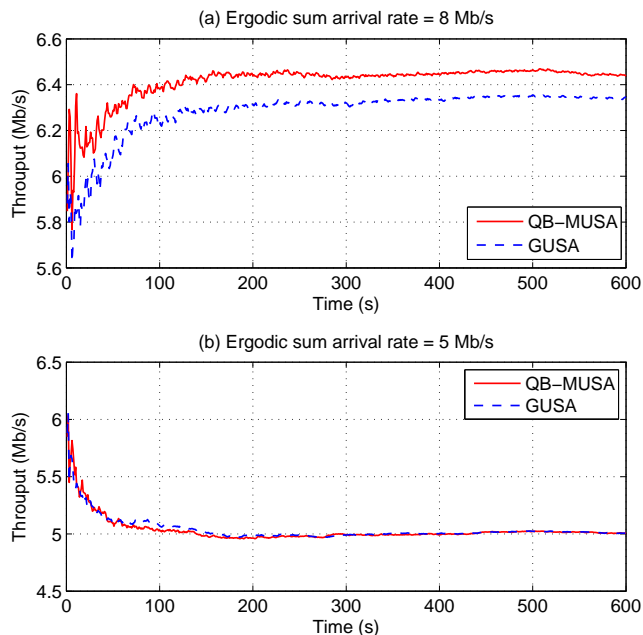
Fig. 7: Throughput comparisons between QB-MUSA and GUSA

index to judge the issue and Fig. 6 indicates the results with basic simulation settings the same as above. Although in our proposed algorithm, users are with different frequencies, the data-rate fairness index of T2-MUSA is higher than FFUSA especially in high SNR regime. RRF user schedules treat every user equally in whatever cases, hence it lies at a high fairness level with a straight flat line. The ISC scheme attempts to maximize and equalize each user's date-rate, if we set the target individual SINRs equal to each other. So ideally the Jain's fairness index of ISC schedules should be inherently 1, meaning that the schedule is absolutely fair. However, due to the use of approximated SINR, users' data-rates derived from real SINR disperse a little. Even so, it makes the highest fairness score among the tested algorithms.

E. Performance Comparison between QB-MUSA and GUSA

In Fig. 7 we simulate 10-minute run of a broadcast system containing 15 users with transmit correlation matrices generated the same as above. The BS is equipped with 8 antennas and the transmit power is fixed at 15 dB. The initial queue lengths are given by random numbers that follow uniform distribution $U(0, 6)$ (with unit in Mb). The arrival process of user k is expressed as

$$A_k(t) = \omega_k A m(t), \tag{46}$$

where ω_k is the fractional arrival rate subject to $\sum_{k=1}^{K} \omega_k = 1$; A is the ergodic sum arrival rate and $m(t)$ stands for the random arrival instance such that $m(t) \in U(0, 2)$. We make the time slot interval equal to 1 second and the bandwidth equal to 1 MHz. The sub-figure (a) demonstrates the simulation results of $A = 8$ Mb/s and (b) $A = 5$ Mb/s. We actually select these two ergodic sum arrival rates to show two different scenarios i.e. (a) A is beyond the system transmission ability and (b)

is on the contrary. In the first case, QB-MUSA realize higher throughput than GUSA, while in the later case both algorithms are able to finish transmitting all queueing data. Hence, by QB-MUSA we see an alternative scheduling algorithm that has the possibility to outperform conventional WSR-based scheduling algorithms.

VI. CONCLUSION

In this paper, we studied two problems in maximizing the spectral efficiency of MU-MISO broadcast channels. Each problem has its own constraints and for each of them we proposed a user scheduling algorithm named T2-MUSA and QB-MUSA. In multi-user system, there is a well-known trade-off between the sum-rate maximization and user fairness. Compared with conventional user scheduling algorithms, the merit of T2-MUSA should be seen as the joint advantage in system throughput and user fairness. Besides, we discussed the optimality of T2-MUSA over higher tier algorithms, and found an analytical expression of the lower bound of the ergodic average group-rate when the dominant eigenvalues of arbitrary users are uniformly distributed. QB-MUSA is designed to deal with the queue stability problem which actually is converted to a throughput maximization problem in our study. The result shows that QB-MUSA is possible to gain higher throughput especially in the case that the ergodic sum arrival rate is larger than the system capacity.

APPENDIX A
PROOF OF THEOREM 1

Suppose not: the optimal fair schedule $S^\star$ is composed of both SU and TU groups.

Without loss of generality, we denote one of the users in SU groups as user k and one of the users in TU groups as user j. Assume there are t_{k1} SU groups and t_{k2} TU groups containing user k and the similar goes for t_{j1} and t_{j2}, where $t_{k1}, t_{j2} \in \mathbb{Z}^+$ and $t_{k2}, t_{j1} \in \mathbb{N}$. Since we have $S^\star \in \mathbb{S}_{FF}$, we can write

$$t_{k1} + t_{k2} = t_{j1} + t_{j2}. \tag{47}$$

Also, because of $S \in \mathbb{S}_{EF}$, each user consumes the same energy i.e.

$$t_{k1}P + t_{k2}\frac{P}{2} = t_{j1}P + t_{j2}\frac{P}{2}. \tag{48}$$

Combining (47) and (48), we can get

$$\begin{cases} t_{k1} = t_{j1} = t_1 \\ t_{k2} = t_{j2} = t_2, \end{cases} \tag{49}$$

which means the number of SU groups that contains user k is equal to the number of SU groups that contains user j and so happens to the number of TU groups. Since k could stand for any user in SU groups and j could represent any user in TU groups, it holds true that every user is allocated to t_1 SU groups and t_2 TU groups. Therefore $S^\star$ can be separated into two set,

$$S^\star = \{S_1, S_2\} \tag{50}$$

where S_1 only consists of SU groups and S_2 only consists of TU groups. Due to the fact $t_1, t_2 \in \mathbb{Z}^+$, S_1 and S_2 are two

complete user schedules. Assume the sizes of S_1 and S_2 are respectively L_1 and L_2 and the average group rates are $\bar{R}_{S_1}$ and $\bar{R}_{S_2}$, then the sum rate of S_1 and S_2 should be equal to the sum rate of $S^\star$,

$$L_1 \bar{R}_{S_1} + L_2 \bar{R}_{S_2} = (L_1 + L_2) \bar{R}_{S^\star} \qquad (51)$$

i.e.

$$\frac{L_1}{L_2} \left(\bar{R}_{S^\star} - \bar{R}_{S_1} \right) = - \left(\bar{R}_{S^\star} - \bar{R}_{S_2} \right). \qquad (52)$$

- If $\bar{R}_{S_1} = \bar{R}_{S_2}$, by (52) we can deduce $\bar{R}_{S_1} = \bar{R}_{S_2} = \bar{R}_{S^\star}$ which contradicts with that $S^\star$ is optimal, since S_1 and S_2 achieve equivalent average group rate with shorter lengths.
- If $\bar{R}_{S_1} \neq \bar{R}_{S_2}$, by (52) we have either $\bar{R}_{S_1} > \bar{R}_{S^\star}$ or $\bar{R}_{S_2} > \bar{R}_{S^\star}$ which also contradicts with that $S^\star$ is optimal.

Therefore, we complete the proof.

APPENDIX B
PROOF OF LEMMA 2

Suppose not: there are only n independent '1's in $\mathbf{B}$ where n is smaller than K.

According to Remark 2 we can find n lines that contain all '1's in $\mathbf{B}$. Without loss of generality, we assume these lines cover the $\{r_1, r_2, \cdots, r_{n_r}\}^{th}$ rows and the $\{c_1, c_2, \cdots, c_{n_c}\}^{th}$ columns with $n_r + n_c = n$. Then we make two sets of lines by

$$\mathcal{I}_\mathcal{R} = \underbrace{\{r_1, r_2, \cdots, r_{n_r}, 2r_1, 2r_2, \cdots, 2r_{n_r}, \cdots, mr_1, mr_2, \cdots, mr_{n_r}\}}_{mn_r \text{ rows}},$$

$$\qquad (53)$$

$$\mathcal{I}_\mathcal{C} = \underbrace{\{c_1, c_2, \cdots, c_{n_c}, 2c_1, 2c_2, \cdots, 2c_{n_c}, \cdots, mc_1, mc_2, \cdots, mc_{n_c}\}}_{mn_c \text{ columns}}.$$

$$\qquad (54)$$

Therefore, we have $m(n_r + n_c) = mn$ lines in total and we can use these mn lines to cover all the '1' elements in $\mathbf{A}_m$. Specifically, if a '1' element is covered by the r_i^{th} row in $\mathbf{B}$, then its associated '1' element in $\mathbf{A}_m$ must be covered by one of the $\{r_i, 2r_i, \cdots, mr_i\}^{th}$ rows . It means there are only mn independent '1's in $\mathbf{A}_m$ which contradicts with that $\mathbf{A}_m$ has mK independent '1's. Hence we complete the proof of Lemma 2.

APPENDIX C
PROOF OF THEOREM 2

We assume $S^\star_{EFT2}$ is the optimal tier-2 energy-fair schedule and $S^\star_{EFTN}$ is the optimal tier-N energy-fair schedule.

- If N is an even number, according to Corolarry 1, we can manage to separate $S^\star_{EFTN}$ as (34),

$$S^\star_{EFTN} = \{S^1_{EFT2}, S^2_{EFT2}, \cdots, S^{\frac{N}{2}}_{EFT2}\}. \qquad (55)$$

While, we also know that among all tier-2 energy-fair schedules, $S^\star_{EFT2}$ is the optimal one i.e.

$$\bar{R}_{S^\star_{EFT2}} \geq \bar{R}_{S^n_{EFT2}}, \quad n = 1, 2, \cdots, N/2. \qquad (56)$$

In other words, the average group rate of $S^\star_{EFT2}$ should be no less than other tier-2 energy fair schedules and we consequently have

$$\bar{R}_{S^\star_{EFT2}} \geq \frac{2}{N} \sum_{n=1}^{N/2} \bar{R}_{S^n_{EFT2}} = \bar{R}_{S^\star_{EFT2m}}. \qquad (57)$$

Combining (30) and (57) gives

$$\bar{R}_{S^\star_{EFT2}} = \bar{R}_{S^\star_{EFTN}}. \qquad (58)$$

- If N is an odd number, considering Lemma 1 and what we have proved in (58), we can write,

$$\bar{R}_{S^\star_{EFT2}} = \bar{R}_{S^\star_{EFT2N}} \geq \bar{R}_{S^\star_{EFTN}} \qquad (59)$$

When there are even number of users, we may be able to find the average group-rate of the tier-1 user schedule $S^\star_{EFT1}$ is equal to $S^\star_{EFT2}$ (in $S^\star_{EFT1}$, each user is allocated to a TU group, so that the energy consumed by an individual user is $P/2$). Only in this case the optimal energy-fair schedule is $S^\star_{EFT1}$, hence we make a condition upon the Theorem 2 i.e. $N \geq 2$.

To sum up, for any tier-N optimal energy-fair schedules, we proved that the average group-rate of $S^\star_{EFT2}$ is no less than the rest while its length is the shortest. Hence the proof of Theorem 2 is completed.

APPENDIX D
PROOF OF PROPOSITION 3

Because the user schedule with fully SU groups (24) belongs to tier-2 schedules i.e. $S^\star_1 \in S_{EFT2}$, the average group-rate of $S^\star_{EFT2}$ should be inherently no less than $S^\star_1$. Suppose the $\lambda \in U(a, b)$ is the maximum eigenvalue of a single user K in $S^\star_1$, by (7), (8) and (9), we have

$$\bar{R}_{S^\star_1} = E[R_K] = E[h(P\lambda)]. \qquad (60)$$

On the other hand we have following two expansions:

$$\lim_{x \to \infty} \exp\left(\frac{1}{x}\right) = \sum_{n=0}^{\infty} \frac{1}{x^n n!} > 1 + \frac{1}{x}, \qquad (61)$$

$$E_1\left(\frac{1}{x}\right) = -\gamma - \log\frac{1}{x} - \sum_{n=1}^{\infty} \frac{1}{nn!}\left(-\frac{1}{x}\right)^n$$
$$\underset{x \to \infty}{>} \frac{1}{x} - \frac{1}{4x^2} - \gamma + \log x. \qquad (62)$$

Specifically, (61) is obtained by Taylor expansion at the point $\frac{1}{x} \to 0$ and (62) is obtained with the help of the asymptotic expansion of [32, 5.1.11, p. 229]. Then we can write,

$$E[h(P\lambda)] = \int_a^b h(P\lambda) f_\Lambda(\lambda) d\lambda$$
$$\overset{*}{>} \frac{1}{P} \int_{aP}^{bP} \left(1 + \frac{1}{x}\right) \left(\frac{1}{x} - \frac{1}{4x^2} - \gamma + \log x\right) \frac{1}{b-a} dx$$
$$= \frac{1}{P} \left(\frac{\log^2 x}{2} + (x + 1 - \gamma)\log x - (\gamma + 1)x\right)\Bigg|_{aP}^{bP}$$

$$\qquad (63)$$

where $f_\Lambda(\lambda)$ denotes the probability density function of λ and $(*)$ is obtained by replacing $x = P\lambda$. Thereby, we obtain the equations (41) and (42).

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[3] J.-W. Lee, J.-Y. Ko, and Y.-H. Lee, "Effect of transmit correlation on the sum-rate capacity of two-user broadcast channels," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2597–2599, 2009.

[4] J. Wang, S. Jin, X. Gao, K.-K. Wong, and E. Au, "Statistical eigenmode-based SDMA for two-user downlink," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5371–5383, Oct 2012.

[5] V. Raghavan, S. V. Hanly, and V. V. Veeravalli, "Statistical beamforming on the grassmann manifold for the two-user broadcast channel," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6464–6489, 2013.

[6] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 528–541, March 2006.

[7] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, 2005.

[8] J. Oh, S.-J. Kim, R. Narasimhan, and J. M. Cioffi, "Transmit power optimization for gaussian vector broadcast channels," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, vol. 4, pp. 2712–2716, IEEE, 2005.

[9] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, pp. 5045–5060, Nov 2006.

[10] H. Joudeh and B. Clerckx, "AMMSE optimization for multiuser MISO systems with imperfect CSIT and perfect CSIR," in *2014 IEEE Global Communications Conference*, pp. 3308–3313, IEEE, 2014.

[11] J. P. González-Coma, M. Joham, P. M. Castro, and L. Castedo, "Qos constrained power minimization in the MISO broadcast channel with imperfect CSI," *Signal Processing*, vol. 131, p. 447455, 2017.

[12] C. Wang and R. D. Murch, "Adaptive downlink multi-user MIMO wireless systems for correlated channels with imperfect CSI," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2435–2446, 2006.

[13] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *CoRR*, vol. abs/1010.1499, 2010.

[14] X. Yi and D. Gesbert, "Precoding methods for the MISO broadcast channel with delayed CSIT," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 1–11, May 2013.

[15] J. Wang, M. Matthaiou, S. Jin, and X. Gao, "Precoder design for multiuser MISO systems exploiting statistical and outdated CSIT," *IEEE Transactions on Communications*, vol. 61, pp. 4551–4564, November 2013.

[16] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, 2002.

[17] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 90–100, 2006.

[18] A. Gründinger, M. Joham, and W. Utschick, "Stochastic transceiver design in multi-antenna channels with statistical channel state information," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3420–3423, IEEE, 2011.

[19] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Semiunitary precoding for spatially correlated MIMO channels," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1284–1298, 2011.

[20] B. Clerckx, G. Kim, and S. Kim, "MU-MIMO with channel statistics-based codebooks in spatially correlated channels," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pp. 1–5, IEEE, 2008.

[21] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.

[22] M. K. Member and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1468–1477, 2007.

[23] Y. Wu, S. Jin, X. Gao, M. R. McKay, and C. Xiao, "Transmit designs for the MIMO broadcast channel with statistical CSI," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4451–4466, 2014.

[24] S. Qiang, S. Chen, J. Shi, Y. Zhang, and G. Xiqi, "User scheduling algorithms for downlink MU-MIMO system based on the SCSI," *IEICE transactions on communications*, vol. 96, no. 2, pp. 651–655, 2013.

[25] Q. Sun, Y. Xue, Y. Zhang, S. Jin, X. Li, and X. Gao, "User assignment for MU-MIMO downlink system based on munkres algorithm," in *2013 International Conference on Wireless Communications & Signal Processing (WCSP)*, pp. 1–6, IEEE, 2013.

[26] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.

[27] F. Bourgeois and J.-C. Lassalle, "An extension of the munkres algorithm for the assignment problem to rectangular matrices," *Communications of the ACM*, vol. 14, no. 12, pp. 802–804, 1971.

[28] Z. Galil, "Efficient algorithms for finding maximum matching in graphs," *ACM Computing Surveys (CSUR)*, vol. 18, no. 1, pp. 23–38, 1986.

[29] S. G. Srinivasan and M. K. Varanasi, "Optimal spatial correlations for the noncoherentMIMO rayleigh fading channel," *IEEE Transactions on Wireless Communications*, vol. 6, no. 10, pp. 3760–3769, 2007.

[30] M. Dai and B. Clerckx, "Transmit beamforming for MISO broadcast channels with statistical and delayed CSIT," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1202–1215, 2015.

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

[32] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1964.

Qi Cao received the B.S. degree in Electronic Communication Engineering from University of Liverpool in 2013, the M.S. degree in Communications Signal Processing from Imperial College. He is now working towards the Ph.D degree in School of Information and Control Engineering of China University of Mining and Technology. His research interests include MIMO wireless communications, Multi-user transmission and full-duplex and cooperative communication.

Yanjing Sun is currently a professor and vice dean of the School of Information and Control Engineering, China University of Mining and Technology. He received the Ph.D. degree in Information and Communication Engineering from China University of Mining and Technology in 2008. His current research interests include IBFD communication, Embedded real-time system, Wireless sensor networks, Cyber-physical system and so on.

Qiang Ni received the B.Sc., M.Sc., and Ph.D. degrees from the Huazhong University of Science and Technology, China, all in engineering. He is currently a Professor and the Head of the Communication Systems Group with InfoLab21, School of Computing and Communications, Lancaster University, Lancaster, U.K. He has authored over 180 papers in international journals and conferences. His main research interests lie in the area of future generation communications and networking, including green communications and networking, cognitive radio network systems, heterogeneous networks, small cell and ultra-dense networks, 5G, SDN, cloud networks, energy harvesting, wireless information and power transfer, IoTs, and vehicular networks. He was an IEEE 802.11 Wireless Standard Working Group Voting Member and a Contributor to the IEEE Wireless Standards.

Zefu Tan received M.S. degree in electromagnetic field and microwave technology from Beijing University of Posts and Telecommunications. He is currently pursuing the Ph.D. degree at School of Information and Electrical Engineering, China University of Mining and Technology. His research interests include wireless communication and signal processing.

Song Li is an associate professor in School of Information and Control Engineering, China University of Mining and Technology. He received the Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications in 2012. His current research interests are full duplex communication, cyber-physical system and cooperative communication.