Performance Analysis of Cooperative Virtual MIMO in Small Cell Networks

Kan Zheng, *Senior Member, IEEE*[*], Xuemei Xin[*], Fei Liu[*], Wei Xiang, *Senior Member, IEEE*[†], and Mischa Dohler *Senior Member, IEEE*[‡]

[*] Key Lab of Universal Wireless Communications, Ministry of Education
Beijing University of Posts & Telecommunications
Beijing, 100088, China
[†] Faculty of Engineering and Surveying
University of Southern Queensland
Toowoomba, QLD 4350, Australia
[‡] Centre Tecnologic de Telecommunications de Catalunya (CTTC)
Barcelona, Spain, 08860

Abstract

With the advent of small cell networks (SCNs) to support growing wireless data volumes and thus reduced cell sizes, cooperative communications are significantly facilitated. Applicable to 3GPP LTE-A, we propose a novel channel/queue-aware user pairing and scheduling scheme in a cooperative virtual multiple-input multiple-output (VMIMO) system. The queueing performance of the VMIMO system with the scheduling scheme is analyzed based on the finite-state Markov model (FSMM), and compared with that of non-cooperative systems. Bounds on the average queuing delay of users are derived by using a semi-definite programming (SDP) approach. The presented analyses are validated through comparing the analytical and simulation results. It is found that the introduced VMIMO pairing process is able to significantly reduce service delays, bringing on a positive impact of cooperative techniques on next generation wireless systems.

Index Terms – VMIMO, Pairing, Finite-State Markov Model (FSMM).

I. INTRODUCTION

It is a challenging task for traditional cellular networks to meet increasing requirements on ubiquitous wireless broadband coverage with high data rates. It is well known that cell-size reduction is the simplest and most effective way to increase network capacity [1]. Towards this end, a new network design concept, known as "small-cell networks (SCNs)", is currently being investigated in the 3rd Generation Partnership Project (3GPP) [2], which can provide cost-effective and energy-efficient solutions to future explosive traffic growth. SCNs are based on the idea of a very dense deployment of self-organizing, low-cost, low-power BSs. The idea of small cells is not entirely new. For example, these so-called small cells are already complemented by picocells for coverage extension and local capacity enhancement. Moreover, the recent launch of femtocells can be seen as the first step toward an unplanned deployment of self-organizing SCNs. However, although as a promising concept, SCNs also lead to many new challenges and opportunities to system design [3]. With a smaller cell size and closer proximity among user equipments (UEs), some form of cooperation between UEs becomes more convenient and frequent.

Virtual multiple-input and multiple-output (VMIMO), as one form of cooperative communications among UEs in uplink transmission, has been widely studied in the 3rd generation (3G) long-term evolution (LTE) and other systems [4] [5]. Also, it is applicable in SCNs of LTE and its advanced embodiment (LTE-A). Due to size limitation, it is difficult to deploy multiple transmit antennas in a portable UE, which prevents spatial multiplexing from being directly applied in the uplink. By virtue of cooperative VMIMO techniques, two or more UEs equipped with only a single antenna can be paired to transmit independent data streams on the same frequency-time resource block (RB) simultaneously. Compared to traditional MIMO, multi-user diversity is achieved in VMIMO through the user pairing and scheduling process. How to select partner users to form a virtual MIMO is crucial for the system performance. Several pairing and scheduling schemes have been proposed in recent years, which tend to strike a good trade-off between throughput and fairness. In [6] and [7], the pairing algorithms based on the determinant of the equivalent channel matrix are studied, which only consider the orthogonality of the channels between users. Then, the signal-to-interference-plus-noise ratio (SINR)-based paring scheme is proposed [8], where two UEs with a small difference in SINR are more likely to be paired. However, these schemes ignore user fairness and are not optimal from the perspective of throughput maximization. In order to improve the fairness performance, single proportional-fairness (SPF) and double PF (DSF) schemes are proposed in [9] and [10] at the expense of throughput loss. Furthermore, joint pairing and resource allocation schemes are studied in [11] and [12], which can achieve a high throughput at the expense of computational complexity.

To the best of our knowledge, there has no pairing *and* scheduling scheme proposed to date in the literature for VMIMO systems *considering* the effects of the queue status. Existing schemes only take into account channel conditions but ignore the queue status, which inevitably causes the waste of wireless resources. Resources are allocated to the users with best channel conditions that may not always have data to transmit, and thus the allocated resources may not be truly utilized. On the other hand, other users whose queues are not empty cannot exploit the resources. Therefore, in order to solve this problem, the scheduler has to take into account not only the channel conditions but the queue states. In other words, only the users whose queue is not empty are allowed to compete for the resources, while others are not allocated any resource even under a good channel condition. Towards this end, we propose a channel/queue-aware scheduling scheme, which selects the user pair and allocates resources based upon not only the channel state but the queue state of the users. In our proposed scheme, the first user is selected among the users with a non-empty queue using the round-robin (RR) criterion. In the second step, this user can transmit its data with or without forming a VMIMO system, which is determined by the scheduling in accordance with the channel environments and queue states. For the sake of analysis, the objective of maximizing the channel capacity is chosen for the second step.

Previous pairing and scheduling schemes focus only on the throughput performance. Due to the dynamic queue status of the system with finite buffer services, the queueing performance becomes very important as well as the throughput performance. However, since the pairing process may affect the service rate of the system, the analysis of the queueing performance of a VMIMO system is fairly difficult, which has not yet been investigated in the literature. The finite-state Markov channel (FSMC) model has been widely used for modeling wireless channels [13] [14]. A wide range of FSMC applications in performance evaluation of wireless networks can be found in the literature, e.g., [15] [16]. This paper adopts the finite-state Markov model (FSMM) to reflect the service rate of the single-input multiple-output (SIMO) and VMIMO channels. After calculating the average service rate of users under a specific queue state, bounds on the queue length of each user can be derived by using a semi-definite programming approach.

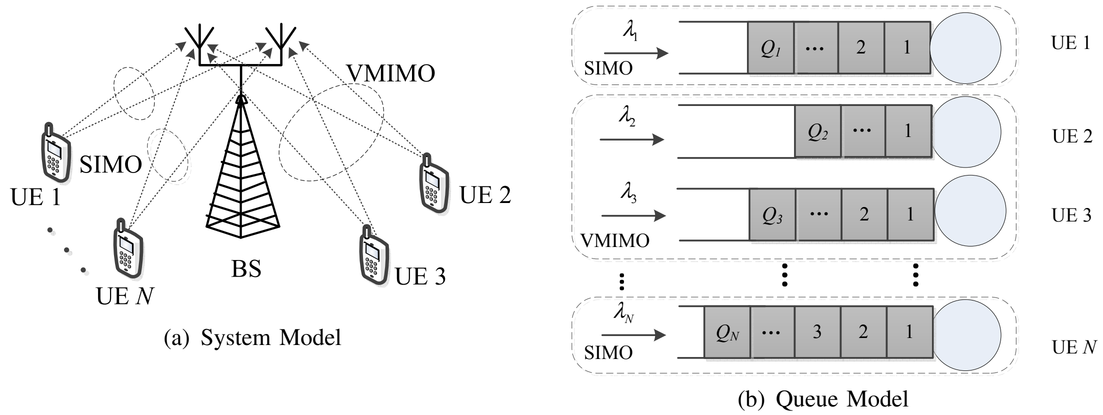


Fig. 1. Illustration of a VMIMO system.

The major contributions of this paper are summarized in the following:

1) A channel/queue-aware pairing and scheduling scheme is proposed, where only the users with a non-empty queue have the opportunity to be scheduled;
2) We propose a method to analyze the queueing performance of a VMIMO system and apply it to study the maximal capacity pairing scheme. To the best of our knowledge, prior to this paper, no good approximation is available for the queueing performance of the VMIMO system; and
3) Analytical bounds for the VMIMO system are derived and validated by means of computer simulations.

The remainder of this paper is organized as follows. In Section II, the flow-level model is formulated for the purpose of analysis. In Section III, a channel/queue pairing and scheduling scheme is first proposed, then the analytical method of the VMIMO system is introduced based on the FSMM model. In Section IV, numerical and simulation results are presented, compared and discussed. Finally, concluding remarks are drawn in Section V.

II. Formulation of Flow-level Model

A. System Model

Consider an uplink system comprising N independent UEs, each having one transmit antenna, and a BS equipped with $N_R \leq N$ receive antennas. The BS is assumed to know the ideal channel state information (CSI) as well as the queue state information of each user. Then, as shown in Fig. 1, through the scheduler at the BS, each UE establishes transmission linkages with the BS either independently or cooperatively through VMIMO, depending on the channel environment and employed scheduling schemes.

If no user is necessary to be paired for transmission, only a single UE, e.g., the nth UE, is scheduled on a given time-frequency resource block. Then, after passing through a $1 \times N_R$ SIMO channel, the received signal at the BS can be expressed as

$$\mathbf{y}'_n = \sqrt{P}_n \rho_n \mathbf{h}_n x_n + \mathbf{z} = [y'_{1,n}, y'_{2,n}, \cdots, y'_{N_R,n}]^T \in \mathbb{C}^{N_R \times 1}, 1 \le n \le N, \tag{1}$$

where x_n represents the transmitted signal from the nth UE with the transmit power of P_n, ρ_n denotes the joint path loss and shadow fading of the nth UE, $\mathbf{h}_n = [h_{1,n}, h_{2,n}, \cdots, h_{N_R,n}]^T \in \mathbb{C}^{N_R \times 1}$ denotes the complex fading channel vector from the nth UE to the BS, and $\mathbf{z} = [z_1, z_2, \cdots, z_{N_R}]^T \in \mathbb{C}^{N_R \times 1}$ is modeled as zero-mean additive white Gaussian noise (AWGN) with the covariance matrix of $\sigma_N \mathbf{I}$. The channel gains between the users' and BS's antennas are assumed to be independent, identically distributed (i.i.d.) Rayleigh fading, i.e., $h_{m,n} \sim \mathcal{CN}(0, \sigma^2)$[1], $1 \le m \le N_R, 1 \le n \le N$, with $\sigma^2 = \mathbb{E}[|h_{m,n}|^2]$. For simplicity, the transmit power is assumed to the same for all the UEs and normalized to one, i.e., $P_n = 1, 1 \le n \le N$.

If the scheduler at the BS chooses K among N users to share the same time-frequency resource blocks, a $K \times N_R$ VMIMO system is constructed. Without loss of generality, the first K UEs are assumed to be paired. Then, the signal received by the BS can be expressed by

$$\mathbf{y} = \sqrt{\mathbf{P}} \mathbf{G} \mathbf{h} \mathbf{x} + \mathbf{z} = [y_1, y_2, \cdots, y_{N_R}]^T \in \mathbb{C}^{N_R \times 1}, \tag{2}$$

where $\mathbf{x} = [x_1, x_2, \cdots, x_K]^T \in \mathbb{C}^{K \times 1}$ represents the transmitted signals from K different users, the diagonal matrix $\mathbf{P} = diag\{P_1, P_2, ..., P_K\} = \mathbf{I}_K$ and $\mathbf{G} = diag\{\rho_1, \rho_2, \cdots, \rho_K\}$ of K-dimension denote the transmit power and the path loss/shadow fading of different users, respectively, and the channel matrix can be represented by

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_K \end{bmatrix} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,K} \\ \vdots & \ddots & \vdots \\ h_{N_R,1} & \cdots & h_{N_R,K} \end{bmatrix} \in \mathbb{C}^{N_R \times K}. \tag{3}$$

At the BS, coherent detection is employed to recover the transmit signal. Different schemes can be used depending on the transmission methods. In the case of SIMO transmission, the signals received at

[1]A circularly symmetric complex Gaussian RV x with mean m and covariance R is denoted by $x \sim \mathcal{CN}(m, R)$.

different antennas as shown in (1) are combined with the maximal ratio combining (MRC) principle as follows

$$\hat{x}_n = \sum_{m=1}^{N_R} \omega_m y'_{m,n} = \sum_{m=1}^{N_R} h^*_{m,n} y'_{m,n} = x_n \rho_n \sum_{m=1}^{N_R} |h_{m,n}|^2 + \sum_{m=1}^{N_R} h^*_{m,n} z_m, \tag{4}$$

where $\omega_m = h^*_{m,n}$ is the MRC weight for the mth branch. Then, the resultant signal-to-noise ratio (SNR) after MRC is simply the sum of the SNRs of each received antenna, i.e.,

$$\gamma_n^{SIMO} = \sum_{m=1}^{N_R} \gamma_{m,n}^{SISO}, \tag{5}$$

where $\gamma_{m,n}^{SISO} = \rho_n^2 |h_{m,n}|^2 / \sigma_N^2$ is the SNR of the mth received antenna.

If VMIMO transmission is adopted, the zero-forcing (ZF) receiver is used to detect the users' signals. The equalized signal is given by

$$\tilde{\mathbf{x}} = \mathbf{w}_{ZF}\mathbf{y} = \mathbf{G}\mathbf{x} + (\mathbf{h}^H\mathbf{h})^{-1}\mathbf{h}^H\mathbf{z}, \tag{6}$$

where $\mathbf{w}_{ZF}$ is the linear detection weight vector defined as

$$\mathbf{w}_{ZF} = (\mathbf{h}^H\mathbf{h})^{-1}\mathbf{h}^H. \tag{7}$$

Thus, the post-processing SNR corresponding to the signal transmitted by the kth antenna, i.e., the antenna of the kth UE, can be calculated as [17]

$$\gamma_k^{\text{VMIMO}} = \frac{\rho_k^2}{\sigma_N^2 (\mathbf{h}^H\mathbf{h})_k^{-1}}, 1 \le k \le K, \tag{8}$$

where $(\mathbf{A})_k^{-1}$ represents the kth diagonal element of the inverse of $\mathbf{A}$.

For illustrative purposes, our study focuses on the scenario with $N_R = 2$, i.e., the BS is equipped with two antennas. Meanwhile, only two users are scheduled on the same resource blocks, i.e., $K = 2$. It is noted that such a configuration is consistent with many practical application scenarios [18].

B. Channel Rate Process Model

The finite-state Markov channel model has been widely adopted as an effective model for characterizing wireless fading channels. By partitioning the range of the capacity into a finite number of intervals, an FSMM for the channel capacity can be constructed as shown in 2. Since the channel capacity achieved by SIMO or VMIMO is different, one has to construct two FSMMs corresponding to both transmission

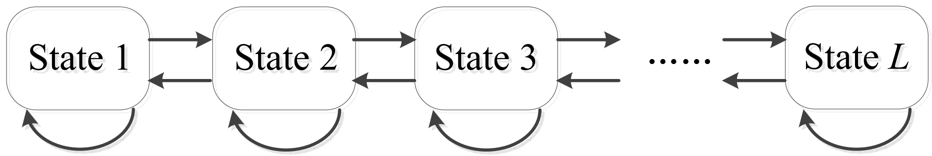


Fig. 2. Illustration of Channel Rate Process Model.

modes. For the implementation of scheduling, the same number of capacity states is used with the same set of capacity thresholds. However, due to the different features of the received SNRs with either SIMO or VMIMO, the capacity state probabilities of the two FSMMs are different, which will be discussed in the following. On the other hand, the transition probability of the FSMM is not needed for our analyses so that they are not presented in this paper.

1) SIMO channel: Under the FSMM, at any time t, the SIMO channel of a user is described by a set of capacity states $\hat{\mathcal{S}} = \{\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_L\}$, where L denotes the number of capacity states of the underlying fading channel. Let Λ_l, $l \in \{1, \ldots, L+1\}$, be the capacity threshold between the lth and $(l+1)$th states of the Markov model for the user. These threshold values are in increasing order with $\Lambda_{L+1} = \infty$. The SIMO channel is in state $\hat{s}_l$ if the instantaneous capacity is between Λ_l and Λ_{l+1}. Corresponding to each state $\hat{s}_l, l \in \{1, 2, \cdots, L\}$, we denote by r_l the service rate of the channel serving workload depending on the wireless channel condition. The rate set, corresponding to L states of the channel, is denoted by $\mathcal{V} = \{r_1, r_2, \cdots, r_L\}$. By convention, the service rate in the worst channel state is set to zero, i.e., $r_1 = 0$.

The stationary probability of the FSMM for the SIMO channel capacity can be defined as follows

$$\hat{\pi}_n^{(l)} = P\{\hat{S}_t = \hat{s}_l\}, 1 \leq n \leq N, 1 \leq l \leq L. \tag{9}$$

According to Shannon's capacity theorem, the normalized capacity of the SIMO channel for the nth user is given by

$$C_n = \log_2\left(1 + \gamma_n^{\text{SIMO}}\right), \tag{10}$$

which is a monotonically increasing function of γ_n^{SIMO}. The probability that the channel capacity state in state $\hat{s}_l$ is determined by the SNR distribution of the SIMO channel, which is given as follows [19]

$$f_S(\gamma) = \frac{1}{\bar{\gamma}^m} \frac{e^{-\gamma/\bar{\gamma}}}{(n-1)!} \gamma^{m-1}, \tag{11}$$

where $\bar{\gamma}$ is the average SNR and $m = \max\{N_R, K\}$. Specially, when $N_R = 2$ and $K = 1$, the stationary probability of the FSMM for the SIMO channel capacity of the nth user can be calculated by

$$\begin{aligned} \hat{\pi}_n^{(l)} &= \int_{\Gamma_l}^{\Gamma_{l+1}} f_S\left(\gamma_n^{\text{SIMO}}\right) d\gamma_n^{\text{SIMO}} \\ &= \left(1 - \frac{\Gamma_{l+1}}{\bar{\gamma}_n}\right) e^{-\frac{\Gamma_{l+1}}{\bar{\gamma}_n}} - \left(1 - \frac{\Gamma_l}{\bar{\gamma}_n}\right) e^{-\frac{\Gamma_l}{\bar{\gamma}_n}}, \end{aligned} \tag{12}$$

where $\Gamma_l = 2^{\Lambda_l} - 1$ is the SNR threshold corresponding to the capacity threshold for the state partition.

2) VMIMO channel: Besides the SIMO channel, an FSMM is constructed for the VMIMO channel capacity when users are paired for transmission, i.e., $\mathcal{S} = \{s_1, s_2, \cdots, s_L\}$ with the same capacity threshold of Λ_l, $l \in \{1, \ldots, L+1\}$ and the rate set of $\mathcal{V} = \{r_1, r_2, \cdots, r_L\}$.

The associated SNR γ_k^{VMIMO} has been shown to be a chi-square random variable with $2(N_R - K + 1)$ degrees of freedom [20]. The cumulative distribution function (CDF) of $\gamma_k^{\text{VMIMO}} \sim \chi_2\left(N_R - K + 1\right)$, with variance 1/2 for the participating Gaussian random variables, is

$$F\left(\gamma_k^{\text{VMIMO}}\right) = 1 - e^{-\gamma_k^{\text{VMIMO}}/\bar{\gamma}_k} \sum_{i=1}^{N_R - K + 1} \frac{\left(\gamma_k^{\text{VMIMO}}/\bar{\gamma}_k\right)^{i-1}}{(i-1)!}. \tag{13}$$

Specially, when two UEs, i.e., UE i and UE j, are paired to form a VMIMO system with two received antennas at the BS, i.e., $N_R = K = 2$, the probability distribution function of γ_k^{VMIMO} can be derived as

$$g\left(\gamma_k^{\text{VMIMO}}\right) = \frac{1}{\bar{\gamma}_k} \exp\left(-\frac{\gamma_k^{\text{VMIMO}}}{\bar{\gamma}_k}\right), k = i,\ j. \tag{14}$$

The BS can calculate the corresponding VMIMO channel capacity as

$$C_{i,j} = \log_2(1 + \gamma_i^{\text{VMIMO}}) + \log_2(1 + \gamma_j^{\text{VMIMO}}). \tag{15}$$

Then, its CDF can be computed by

$$f_V\left(c\right) = \frac{1}{\bar{\gamma}_i \bar{\gamma}_j} \left(\ln 2\right)^2 2^c e^{\frac{1}{\bar{\gamma}_i} + \frac{1}{\bar{\gamma}_j}} \int\limits_0^c e^{\frac{1}{\bar{\gamma}_i} 2^x + \frac{1}{\bar{\gamma}_j} 2^{c-x}} dx. \tag{16}$$

The stationary probability of the FSMM for the VMIMO channel capacity can be computed as

$$\pi_{i,j}^{(l)} = P\left\{S_t = s_l\right\} = \int\limits_{\Lambda_l}^{\Lambda_{l+1}} f_V\left(c\right) dc = \frac{1}{\bar{\gamma}_i \bar{\gamma}_j} \left(\ln 2\right)^2 e^{\frac{1}{\bar{\gamma}_i} + \frac{1}{\bar{\gamma}_j}} \int\limits_{\Lambda_l}^{\Lambda_{l+1}} 2^c \int\limits_0^c e^{\frac{1}{\bar{\gamma}_i} 2^x + \frac{1}{\bar{\gamma}_j} 2^{c-x}} dx dc. \tag{17}$$

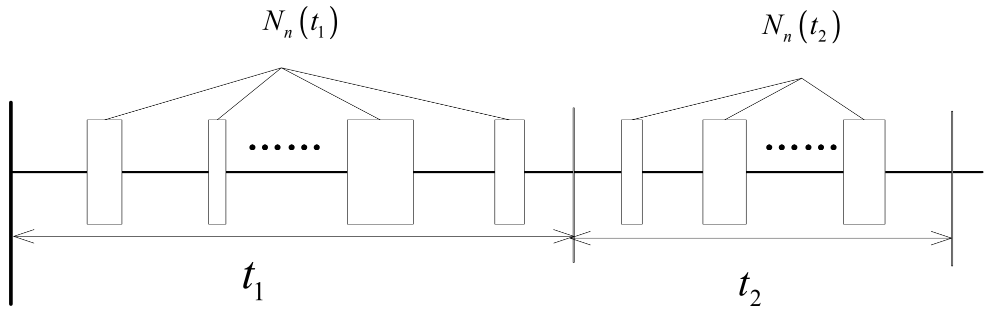


Fig. 3. Illustration of the Dynamic Flow with the Poisson distribution.

C. Dynamic Flow Model

In this paper, we focus on the performance at the flow level in a dynamic setting with random finite-size service demands. A dynamic flow model with elastic traffic is assumed, where a new flow arrives at the system with a finite-length file request, and leaves the system after the file is transmitted. As shown in Fig. 3, the flow of user n arriving at the network follows a Poisson distribution with an average arrival rate of λ_n. The number of arrivals $N_n(t)$ of user n in a finite interval of length t obeys the Poisson distribution, i.e.,

$$P\left\{N_n\left(t\right)=m\right\}=\frac{\left(\lambda_n t\right)^m}{m!}e^{-\lambda_n t}. \tag{18}$$

The exponential flow sizes are independent and identically distributed, which can be expressed as

$$P\left\{F_n \le a\right\}=1-e^{-a/\bar{F}_n}, \tag{19}$$

where F_n is the file length of user n, and $\bar{F}_n=\mathrm{E}[F_n]$ is the mean size of the flow for user n. The flows are served with a first-come first-serve (FCFS) policy at each link. Each UE is assumed to start a new transmission only after the old one is finished, and each new transmission by the same UE is treated as a new flow. After a flow is sent over, it is cleared off from the user queue, and the queue length is decreased by one. Let $\mathbf{Q}\left(\tau\right)=\left\{Q_n\left(\tau\right),n=1,2,...,N\right\}$ be the queue length of every user at time τ, and $\mathbf{\Theta}\left(\tau\right)=\left\{\Theta_n\left(\tau\right),n=1,2,...,N\right\}$ denotes the queue status, where $\Theta_n\left(\tau\right)$ can take only on values of 0 or 1 with $\Theta_n\left(\tau\right)=1$ indicating that the queue of user n is not empty. Note that $\mathbf{\Theta}\left(\tau\right)$ can take 2^N possible values. $Q_n\left(\tau\right)$ as well as $\Theta_n\left(\tau\right)$ depend on both the packet arrival process and the different scheduling strategy at time τ. When user n is scheduled, the queue length of user n is decreased. The average service rates at various queue states depend on the fixed $\mathbf{\Theta}\left(\tau\right)$.

TABLE I
SUMMARY OF MAIN ABBREVIATION AND SYMBOLS

Acronym/Symbol	Definition
VMIMO	Virtual multiple-input multiple-output
SIMO	Single-input multiple-output
AWGN	Additive white Gaussian noise
MRC	The maximal ratio combining
ZF	Zero-forcing
FSMM	the finite-state Markov model
N	The number of UEs in the uplink system
N_R	The number of receive antennas of the BS
K	The number of UEs chosen to construct VMIMO
L	The number of capacity states of underlying fading channel
x_n	The transmitted signals from the nth UE
ρ_n	The joint path loss and shadow fading of the nth UE
P_n	The transmit power of the nth UE
γ_n^{SIMO}	The sum of the SNRs of each received antenna corresponding to the nth UE
$\gamma_{m,n}^{\text{SISO}}$	The SNR of the mth received antenna corresponding to the nth UE
γ_k^{VMIMO}	The post-processing SNR corresponding to the signal transmitted by the antenna of the kth UE
C_n	The normalized capacity of the SIMO channel for the nth UE
$C_{i,j}$	The corresponding VMIMO channel capacity for the ith and the jth UE
$\hat{s}_l$	The lth capacity state of the FSMM model of the SIMO and VMIMO channel
Λ_l	The capacity threshold between the lth and $(l+1)$th states of the FSMM model of the SIMO and VMIMO channel
r_l	The service rate of the channel serving workload depending on the wireless channel
$Q_n(\tau)$	The queue length of the nth UE at time τ
$\Theta_n(\tau)$	The queue status(empty or non-empty) of the nth UE at time τ
$\hat{\pi}_n^{(l)}$	The stationary probability that the SIMO channel capacity state of the nth UE is in the lth state
$\pi_{i,j}^{(l)}$	The stationary probability that the corresponding VMIMO channel capacity state of the ith and the jth UE is in the lth state

As a large number of acronyms and symbols are used in this paper, Table I lists the important ones.

III. A CHANNEL/QUEUE-AWARE SCHEDULING SCHEME

The proposed scheduling scheme can accommodate time-varying channel environments due to fast fading and physical locations of different users, and select proper users in consideration of their channel conditions. Considering the balance between resource efficiency and communication reliability, the scheduler can assign resources to UEs to establish communication either by being paired with each other or by directly communicating with the BS. Besides considering the channel environment, the queue states of users are taken into account too.

A. Maximize Capacity Pairing and Scheduling (MCPS) Scheme

The essential idea is to narrow the scheduling space of the users, allocating resources to the users who have data to transmit rather than all the users in the system. For example, under the specific queue case of $\boldsymbol{\Theta}(\tau)$, when $\sum_{i=1}^{N} \Theta_i(\tau) = N_q(\tau)$, the scheduling decision is only made among the $N_q(\tau)$ users instead of N users. Therefore, the opportunity that the users with a non-empty queue are scheduled is increased.

There are two steps in the scheduling scheme. In the first step, all the RBs are assigned to all the UEs by the RR criterion. In other words, any one of the UEs is selected among the N users by RR as the primary user to occupy one of the RBs. Usually, the number of RBs is larger than that of the users. For example, in the case of 5 MHz bandwidth, the LTE system has 25 RBs, whilst there are usually only $1\tilde{5}$ UEs in a cell in small cell networks. Thus, each UE can be a primary user occupying at least one RB. Without loss of generality, let us take UE 1 as an example for the the purpose of elaboration. In the second step, according to a specific objective or principle, the scheduler decides whether UE 1 solely occupies this RB for SIMO transmission or shares it with a secondary user with a non-empty queue to form a VMIMO system. In other words, as the primary user, UE 1 has two possible ways to transmit its data on this allocated RB, i.e., SIMO and VMIMO. On the other hand, it is possible that UE 1 may be paired with other UEs as a secondary user of a VMIMO system on this RB when other UEs are scheduled with the RR criterion. Different pairing schemes lead to different probabilities of these three types of events.

In the following, we propose the maximizing capacity pairing and scheduling (MCPS) scheme, which always chooses the user or a pair of users with the largest capacity for transmission, or equivalently, the best transmission data rate is guaranteed in every scheduling instant.

At the beginning, the index of the UE that may be paired with UE 1 can be found after exhaustive search, i.e.,

$$j^* = \arg \max_{j=2,3,\cdots,N;\Theta_j(\tau)=1} C_{1,j}. \tag{20}$$

Then, the capacities of SIMO and VMIMO transmissions are compared in order to choose the suitable transmission mode, where the one with larger capacity is selected. For VMIMO transmission, the channel capacity allocated to UE 1 is only a proportion of the overall VMIMO capacity as shown in (15). Otherwise, it equals the capacity of the SIMO channel between UE 1 and the BS, i.e.,

$$\eta_1 = \begin{cases} \rho_{1,j} C_{1,j}, & \text{for VMIMO} \\ C_1, & \text{otherwise} \end{cases} \tag{21}$$

where $\rho_{1,j} = \log_2(1+\gamma_1^{\text{VMIMO}})/\{\log_2(1+\gamma_1^{\text{VMIMO}}) + \log_2(1+\gamma_j^{\text{VMIMO}})\}$.

Evidently, as shown in (20), the dimension of such a user pair subset is $N_q(\tau)$. Consequently, the search complexity can be significantly reduced if the number of UEs to be possibly paired is reduced, e.g., set the SNR threshold before pairing.

B. Performance Analysis

Since the instantaneous transmission rate of a flow varies with time due to channel fading, its delay and throughput performances are difficult to analyze. However, assuming that channel fading is much faster than the flow number variation speed, the rate fluctuation of a flow is "averaged out" during its transmission period, and the flow transmission rate can be approximated by a deterministic value [15]. Moreover, the deterministic transmission rate of a flow varies with the number of flows in the system, due to the scheduling gain of the opportunistic scheduler. It is stated in [15] that if an opportunistic scheduler can ensure fair resource sharing in the OFDM system, its queueing behavior can be analyzed by a processor-sharing (PS) model, where the service rate of each flow varies with the number of flows (states) in the system. Obviously, the queueing model of the RR scheduler can be represented by a PS queue with a constant unit service rate. Then, the system can be formulated as a PS model with the average arrival rate vector $\{\lambda_n, n = 1, \ldots, N\}$. Each flow shares a $1/N_q(\tau)$ fraction of the total service rate at time τ.

Without loss of generality, let us take UE 1 as an example to analyze the the effective throughput performance in a VMIMO system by using the FSMM. The MCPS algorithm may assign the resources to either SIMO or VMIMO transmission under the following three cases:

- *Case 1*: Only when its SIMO channel capacity is larger than its VMIMO counterpart with any of all the possible pairs among UEs with non-empty queues, i.e., $C_1 > C_{1,j}, j = 2, 3, \cdots, N, \Theta_j(\tau) = 1$, UE 1 will not be paired with other UEs but transmits its data solely using the assigned RB;
- *Case 2*: If one or more VMIMO channels have larger capacity than the SIMO channel, i.e., $C_{1,j^*} >$

$C_1, j^* = \arg \max_{j=2,3,\ldots,N,\ \Theta_j(\tau)=1} C_{1,j}$, UE 1 will be paired as the primary user with the UE with the maximum VMIMO channel capacity; and

- *Case 3*: UE 1 may also be paired as a secondary user with UE i as the primary user when the capacity of the VMIMO channel between UE 1 and UE i is larger than that of others, i.e., $C_{i,1} > C_i$ and $C_{i,1} > C_{i,j}, j = 2, 3, \cdots, N, i \neq j, \Theta_j(\tau) = 1$.

Under the condition that UE 1 is in the given channel state of either $\hat{s}_k$ for SIMO or s_k for VMIMO, the probabilities of these three cases can be computed as follows

$$\begin{aligned} \alpha_1^{(k)} &= P\{C_1 > C_{1,j}, j = 2, 3, \cdots, N, \Theta_j(\tau) = 1 | \hat{S} = \hat{s}_k\} \\ &= \prod_{j=2,\Theta_j(\tau)=1}^{N} \left(\sum_{l=1}^{k-1} \pi_{1,j}^{(l)} \right) \end{aligned} \tag{22}$$

$$\begin{aligned} \beta_{1,j}^{(k)} &= P\{C_{1,j^*} > C_1, j^* = \arg \max_{j=2,3,\ldots,N,\ \Theta_j(\tau)=1} C_{1,j} | S = s_k\} \\ &= \left(\sum_{l=1}^{k-1} \pi_1^{(l)} \right) \prod_{n=2,n\neq j,\Theta_n(\tau)=1}^{N} \left(\sum_{l=1}^{k-1} \pi_{1,n}^{(l)} \right) \end{aligned} \tag{23}$$

$$\begin{aligned} \theta_{i,1}^{(k)} &= P\{C_{i,1} > C_i, C_{i,1} > C_{i,j}, j = 2, 3, \cdots, N, i \neq j, \Theta_j(\tau) = 1 | S = s_k\} \\ &= P\{C_{i,1} > C_i | S = s_k\} P\{C_{i,1} > C_{i,j}, j = 2, 3, \cdots, N, i \neq j, \Theta_j(\tau) = 1 | S = s_k\} \\ &= \left(\sum_{l=1}^{k-1} \pi_i^{(l)} \right) \prod_{n=2,n\neq i,\Theta_n(\tau)=1}^{N} \left(\sum_{l=1}^{k-1} \pi_{i,n}^{(l)} \right). \end{aligned} \tag{24}$$

Correspondingly, the transmission rates under the three cases are r_k, $\rho_{1,j} r_k$ and $\rho_{1,i} r_k$, respectively. Therefore, under the specific queue case of $\boldsymbol{\Theta}(\tau)$, the average service rate for UE 1 can be computed by

$$\mu_1^{\boldsymbol{\Theta}(\tau)} = \frac{1}{N_q(\tau)\bar{F}} \sum_{k=1}^{L} (\hat{\pi}_1^{(k)} \alpha_1^{(k)} r_k + \sum_{j=2,\Theta_j(\tau)=1}^{N} \pi_{1,j}^{(k)} \beta_{1,j}^{(k)} \rho_{1,j} r_k + \sum_{i=2,\Theta_i(\tau)=1}^{N} \pi_{i,1}^{(k)} \theta_{i,1}^{(k)} \rho_{1,i} r_k). \tag{25}$$

Then, the average service rate for other UEs, i.e., $\mu_n^{\boldsymbol{\Theta}(\tau)}, n = 2, 3, \cdots, N$, can also be calculated by using this method. The service rate of the queue depends on the subset of queue states in the system with a non-empty length, where the number of possible subsets is 2^N.

Next, bounds of the queue length in the VMIMO system can be obtained by a semi-definite programming-based approach. These bounds can be made progressively tight at the expense of the computational complexity of the associated semi-definite program. The queueing system is assumed to be stable with the maximum rate μ^* that bounds the service rate of any UE. The queue length process can be modeled as a continuous-time Markov chain with the service rate calculated above. The continuous-time Markov chain can be uniformized because it is bounded by $\xi = \sum_{n=1}^{N} \lambda_n + N\mu^*$. The state of the uniformized discrete-time Markov chain at time τ is denoted by $\mathbf{Q}(\tau) = \{Q_n(\tau), n = 1, 2, \cdots, N\}$, where $Q_n(\tau)$ is the queue length of UE n at time τ. As mentioned before, $\mathbf{\Theta}(\tau)$ is the states of the queues of all UEs.

The transition probabilities for the uniformized Markov chain are given below

$$\begin{aligned} P\{\text{Arrival into queue } n\} &= \frac{\lambda_n}{\xi}, n = 1, 2, \cdots, N, \\ P\{\text{Departure from queue } n\} &= \frac{\mu_n^{\mathbf{\Theta}(\tau)} \varepsilon_n}{\xi}, n = 1, 2, \cdots, N, \\ P\{\text{No change in state}\} &= 1 - \frac{\sum_{n=1}^{N} (\lambda_n + \mu_n^{\mathbf{\Theta}(\tau)} \varepsilon_n)}{\xi} \end{aligned} \tag{26}$$

where ε_n is equal to one only if $\Theta_n(\tau) = 1$, or it is zero.

This uniformized Markov chain has the same steady state queue length distribution as the original one. Its evolution can be represented by the following stochastic recursion

$$\mathbf{Q}(k+1) = \mathbf{Q}(k) + \mathbf{X}(k), k = 0, 1, \cdots, \tag{27}$$

where $\mathbf{X}(k) = \{X_n(k), n = 1, 2, \cdots, N\}$ is the increment of the queue length. $X(k) = 1$ represents an arrival into queue n at iteration k, whereas $X(k) = -1$ indicates departure. Moreover, $\mathbf{X}(k) = \mathbf{0}$ if the transition corresponds to a self-loop. It is clear that the distribution of $\mathbf{X}(k)$ is dependent on $\mathbf{\Theta}(k)$ and $\mathbf{Q}(k)$.

After obtaining the transition probabilities of the uniformized Markov chain, the moments of $\mathbf{X}(k)$ can be calculated. Then, the lower and upper bounds of the average queue length of each user, denoted by $\bar{Q}_n$, can be solved by using a semi-definite programming approach developed in [21]. Next, the average queuing delay of each user is derived using Little's Law, i.e.,

$$\bar{\phi}_n = \bar{Q}_n / \lambda_n, n = 1, 2, \cdots, N. \tag{28}$$

Also, the average user throughput can be calculated by

$$\bar{\xi}_n = \bar{F}_n / \bar{\phi}_n, n = 1, 2, \cdots, N. \tag{29}$$

C. Computational Complexity and Overhead

The computational complexity of the proposed scheduling scheme is due mainly to not only SNR estimation but also exhaustive search. According to (5) and (8) for SNR estimation, the addition and multiplication operations can be omitted in comparison with the operations for the multiplication and the inverse of the channel matrix in (8), which depends on the matrix dimension, i.e., $O\left(K^2 N_R\right)$ and $O\left(\min(K, N_R)^3\right)$ respectively. For the exhaustive search shown in (20), the number of comparisons is determined by the number of users with a non-empty queue, i.e., N_q. SNR estimation has to be carried out for each comparison. Therefore, the total computational complexity is $O\left(\min(K, N_R)^3 N_q\right) + O\left(K^2 N_R N_q\right)$.

All the control is done by the network. That is, resource allocation for both uplink and downlink, user paring and transmission mode selection (VMIMO or SIMO) is controlled by the BS. At each transmission time, the eNB informs UE of resource allocation on uplink via the control channel. It is not necessary for a UE to know what the transmission mode is, i.e., VMIMO or SIMO. Since the BS made the scheduling decision, it know how to detect the received signals from UEs by which transmission mode. Therefore, there is no additional signaling overhead necessary for feedback in the case of the scheduling with paring compared to the traditional method without paring.

IV. Numerical and Simulation Results

In this section, the queueing performance of the VMIMO-based multi-user wireless network is evaluated through numerical methods as well as Monte Carlo simulations. Main parameters and configurations of the network in our simulations are listed in Table II. Due to the constraint of computational complexity, only a single cell is considered, where the receive antennas in the BS is assumed to be two and the transmit antennas of the UEs is 1. ZF detection is employed on the uplink of VMIMO transmission.

For simplicity, a periodical source with the same process is assumed for each user. The flows arriving to the network follow a Poisson process with the same average arrival rate, i.e., $\lambda_n, n = 1, 2, \cdots, N$, and the mean size of the flows is set to $\bar{F}_n$= 1 Mbits.

Our simulation program is built up on the MATLAB platform. Each user has its buffer, where the arrival packets wait to be transmitted. At the start of each time unit, the scheduler allocates the radio resources

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	2 GHz
Bandwidth	5 MHz
Time slot	1 ms
Antenna configuration	UE:1 Tx BS:2 Rx
Channel model	Rayleigh fading
UE transmit power	20 dBm for 5 MHz
Arrival process	Poisson
VMIMO detection	ZF

TABLE III
AVERAGE SERVICE RATE UNDER THE GIVEN QUEUE STATES (3 UES)

Queue State			Service Rate (Mbps)		
UE1	UE2	UE3	UE1	UE2	UE3
1	0	0	18.738	0	0
0	1	0	0	23.356	0
0	0	1	0	0	28.14
1	1	0	13.25	18.479	0
0	1	1	12.815	0	23.244
1	0	1	0	17.256	22.539
1	1	1	8.625	13.202	19.147

to users according to the adopted channel sharing scheme. After scheduling, the number of the packets in each user's buffer is counted to analyze the backlog performance. Meanwhile, the sojourn time of each packet in the buffer is recorded when the packet is transmitted. The queueing time of the ith flow in the queue of UE n is denoted by $T_{n,i}$, which is the duration from its arrival at the queue to the departure after being served. Then, the average delay of UE n can be collected as the mean sojourn time of all its flows, i.e. ,

$$\bar{\phi}_n' = \frac{\sum_{i=1}^{N_F} T_{n,i}}{N_F}, \ n = 1, 2, \cdots, N, \tag{30}$$

where N_F is the number of flows with the Poisson distribution arriving at the queue of UE n during simulations.

For the purpose of elaboration, the simple case of 3 UEs is first studied, where the average received SNRs are $\bar{\gamma}_1 = 9$ dB, $\bar{\gamma}_2 = 12$ dB, and $\bar{\gamma}_3 = 15$ dB. The queue-state-dependent service rates of all the UEs are calculated and shown in Table III. Each UE can be served by different rates according to its channel environment only when its queue is not empty. If only one of the UEs has data in its queue, the radio resources are dedicated to the UE with a high service rate with SIMO transmission. When more than one users have a non-empty queue, the resources are shared among them through VMIMO. In these cases,

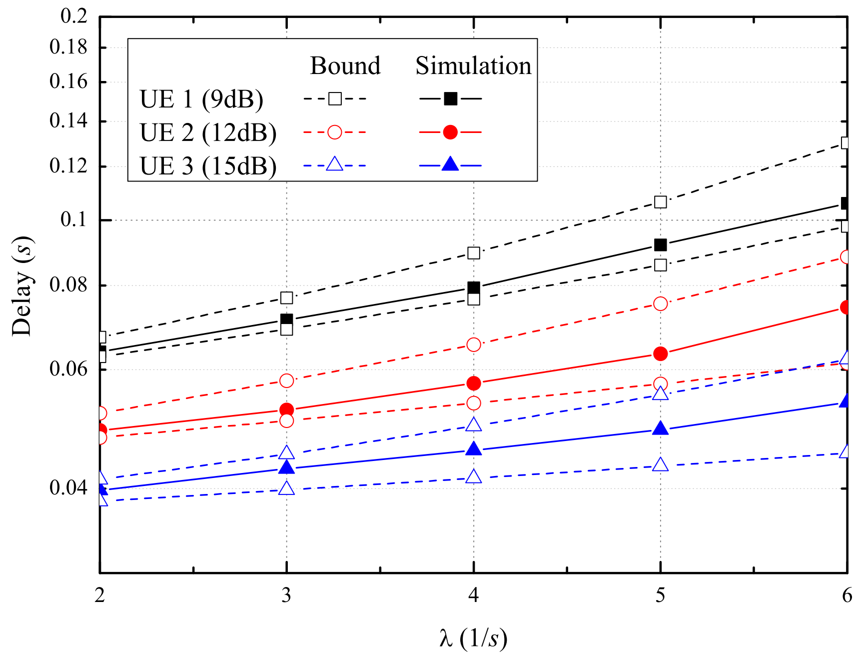


Fig. 4. Comparison of simulation results and analytical bounds for a VMIMO system (3 UEs).

although the service rate of an individual UE is not the highest from its own point of view, the sum of the service rates becomes larger, which is desirable from the system point of view. Fig. 4 compares the analytical bounds and simulation results of a VMIMO with 3 UEs under various average arrival rates, where the average arrival rates for all UEs are assumed to be same. The analytical results are in line with the simulated ones, which demonstrates the effectiveness of the proposed analytical approach. The figure shows that, with the increase of the average arrival rate per user, the delay increases as expected.

Fig. 5 shows the delay performances of three UEs with the same SNR of 12 dB, when varying the average arrival rate for UE 2 and with the given average arrival rate of $4s^{-1}$ for UE 1 and UE 3. It is expected that the delay performance deteriorates with the increase of the average arrival rate, especially for UE 2. Since the network is essentially a queueing system whose the service rate is limited by the wireless channel, the classic queueing theory states that the delay increases with the system load.

Fig. 6 shows the delay and average user throughput performances of the systems with or without pairing, where the number of users varies. The average arrival rate is set to $4s^{-1}$, and the SNRs of all the UEs are set to 12 dB. It is demonstrated that the VMIMO with the proposed pairing and scheduling scheme

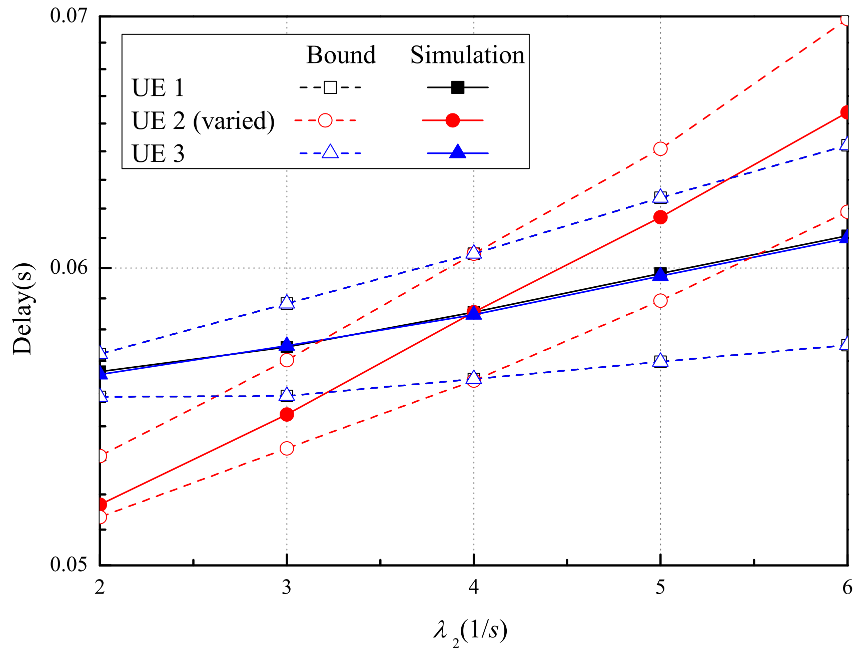
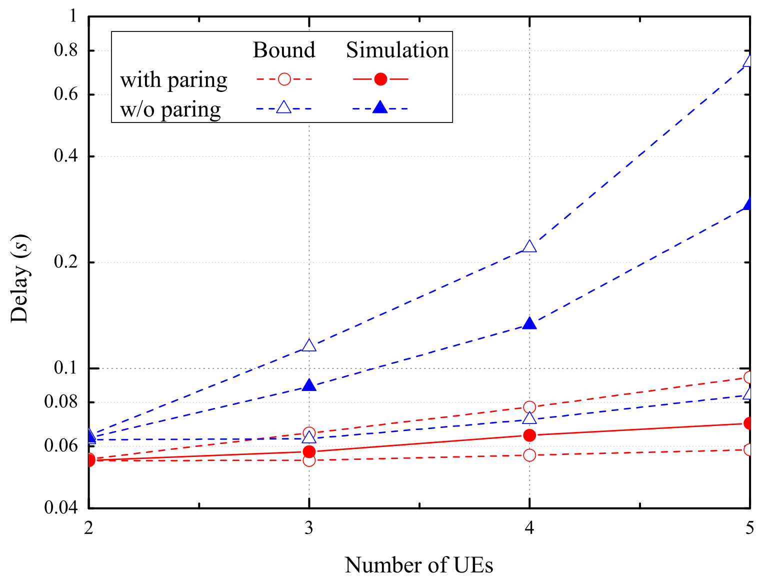


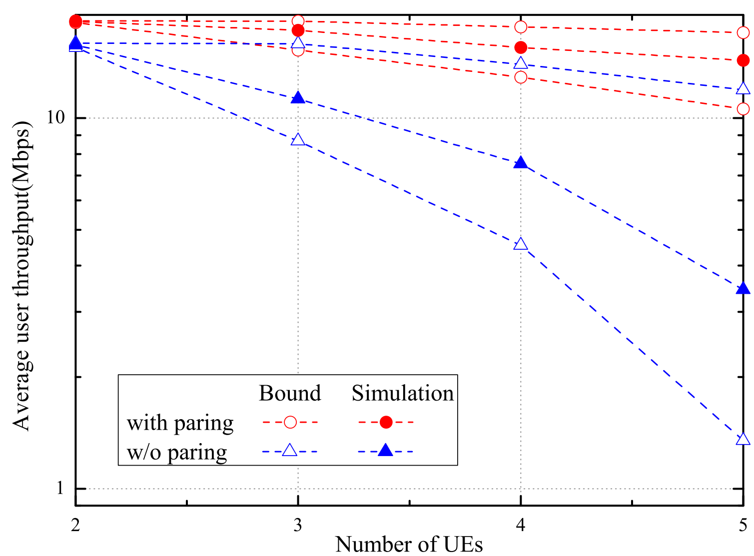
Fig. 5. Delay performances of UEs with the same SNRs (3 UEs, $\lambda_1 = \lambda_3 = 4\ s^{-1}$).

significantly improves the delay performance as well as the average throughput compared with the one without using any pairing scheme. Furthermore, due to the multi-user scheduling gains, the delay and average user throughput performances of the system with paring and scheduling will not significantly degrade with an increase in the number of users. However, those without pairing deteriorate rapidly with more users in the system. Moreover, it is noted that the bound error becomes larger with the number of users, which can be reduced by increasing the moment of the $\mathbf{X}(k)$ in the SDP approach at the expense of computational complexity.

Finally, in order to observe the effect of the channel environment on the delay performance, we only change the SNR of UE 2 while keeping that of the other two UEs unchanged, i.e., $\bar{\gamma}_1 = 9$ dB and $\bar{\gamma}_3 = 15$ dB. The average arrival rate is kept be $4s^{-1}$. As can be seen from Fig. 7, the higher the SNR that UE 2 has, the better its delay performance is. This is attributed to the fact that, when the SNR increases, the channel is more likely (in total probability) to be paired as demonstrated by the stationary probability vectors. This implies that the user is more likely to be served in time, reducing the delay of the system. Meanwhile, the delay performances of the other two UEs are also slightly improved.



(a) Delay



(b) Average user throughput

Fig. 6. Delay and average throughput performances of the systems with/without pairing.

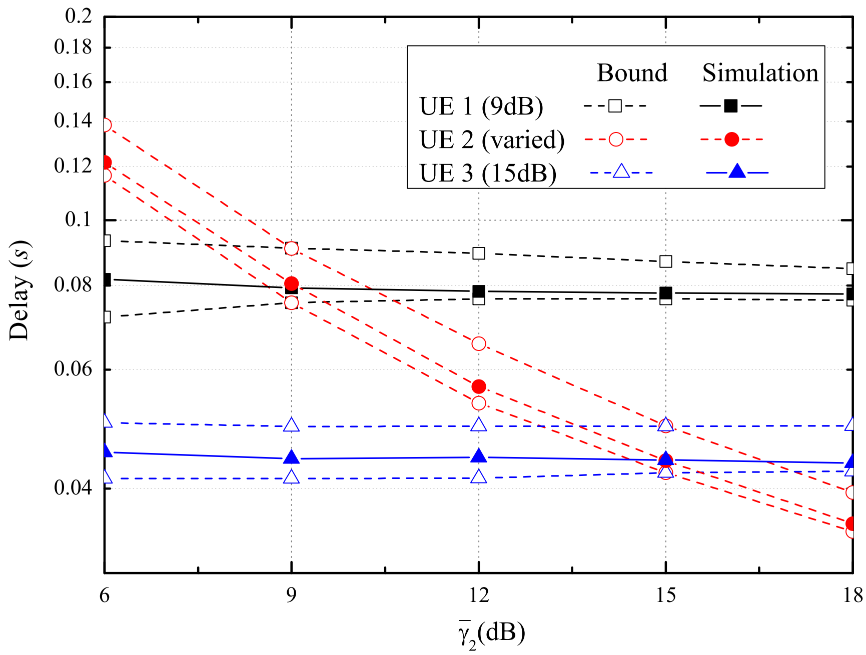


Fig. 7. Delay performance under different SNR in a VMIMO system (3 UEs).

In summary, the performances including delay and throughput can be improved by using VMIMO with the proposed paring schemes. The system performances are dependent on several parameters such as the user number, the SNR difference and the average arrival rate. With the increase of the SNR, the delay reduces while the service rate of the system increases. However, since the system is essentially a queueing system, the average arrival rate should not be more than the service rate limited by the wireless channel. Our numerical results are in line with the simulated ones. However, the bound error becomes larger when the user number is large.

V. Conclusions

In this paper, the analytical methodology for the delay bounds of a VMIMO system with multiple users is developed based on the FSMM. To better exploit limited radio resources, a channel/queue-aware pairing and scheduling scheme was proposed in consideration of the effects of not only the channel environments but also the queue states. A semi-definite programming approach was applied to approximate the average queue length of each user, which leads to the derivation of the bounds of the average queuing delay. Both numerical and simulation results under various system parameter settings were presented and compared, indicating a sufficiently good match between the analytical bounds and the simulation results and thereby validating our presented theoretical analyses. Under the considered scenarios, it was shown that using VMIMO outperforms non-cooperative systems in terms of service delays. Typically, a delay improvement by a factor of two is achievable due to an increasing number of pairable users in the system. We believe that the scheduler design jointly considering channel and queue states allows the use of cooperative VMIMO techniques in forthcoming LTE/LTE-A wireless systems.

By using the proposed methodology, we derived the queue performances of cooperative virtual MIMO in small cell networks without time-consuming simulations. One limitation of our analyses is that the bound error becomes larger with the user number, which has to be dealt with by the more computation-intensive SDP approach. The trade-off between accuracy and computational complexity needs be further investigated. Moreover, in this paper, we have assumed the greedy algorithm, i.e., MCPS, which is optimal only in terms of capacity-approaching but not so under other criteria. For example, the greedy algorithm does not ensure fairness among users. There are many other scheduling algorithms for wireless networks, which may perform better than the greedy algorithm under certain criteria. It will be interesting to investigate how to extend our current analyses to other paring and scheduling algorithms in the future.

Acknowledgment

This work was funded in part by China NSFC (No.61271183), National Basic Research Program of China (973 Program: 2012CB316005), and Program for New Century Excellent Talents in University (NCET-11-0600).

References

[1] Z. Roth, M. Goldhamer, N. Chayat, A. Burr, M. Dohler, N. Bartzoudis, C. Walker, Y. Leibe, C. Oestges, M. Brzozowy, and I. Bucaille, "Vision and Architecture Supporting Wireless GBit/sec/km2 Capacity Density Deployments," Future Network and Mobile Summit 2010, 16-18 June 2010, Florence, Italy.

[2] RWS-120002, NSN, Release 12 and beyond for C^4 (Cost, Coverage, Coordination with small cells and Capacity), June 11-12, 2012.

[3] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37- 43, 2011.

[4] C.S. Park, Y. Wang, G. Jongren, and D. Hammarwall, "Evolution of uplink MIMO for LTE-advanced," *IEEE Communications Magazine*, vol. 49, no. 2, pp.112-121, Feb, 2011.

[5] M. Dohler, E. Lefranc, and A.H. Aghvami, "Virtual Antenna Arrays for Future Wireless Mobile Communication Systems," in Proc. of *IEEE ICT'*2002, Beijing, China, June 2002.

[6] 3GPP TSG-RAN1 WG1 #43, R1-051422, "UL virtual MIMO system level performance evaluation for E-UTRA," Seoul, Korea, Nov. 2005.

[7] B. Fan, W. Wang, Y. Lin, L. Huang, and K. Zheng, "Spatial multi-user pairing for uplink virtual-MIMO systems with linear receiver," in Proc of *IEEE conference on Wireless Communications & Networking Conference*,WCNC'09, pp. 1807-1811, 2009.

[8] J. Han, X. Tao, and Q. Cui, "Simplified SINR-Based User Pairing Scheduling for Virtual MIMO," in Proc. of *IEEE Vehicular Technology Conference*, VTC Spring 2009. pp.1- 4.

[9] 3GPP TSG-RAN1 #46, R1-062052, "UL system analysis with SDMA," Tallinn, Estonia, Aug. 2006.

[10] X. Chen, H. Hu, H. Wang, H. Chen, and M. Guizani, "Double proportional fair user pairing algorithm for uplink virtual MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2425C2429, July 2008.

[11] J. W. Kim, I. S. Hwang, and C. G. Kang, "Scheduling for virtual MIMO in single carrier FDMA (SC-FDMA) system, in Proc. of *Int. Inf. Commun. Technol. Convergence Conf*, pp. 115C119.

[12] J. Fan, G. Li, Q. Yin, B. Peng, and X. Zhu, "Joint User Pairing and Resource Allocation for LTE Uplink Transmission," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2838 - 2847, 2012.

[13] H. Wang and N. Moayeri, " Finite-state Markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol.44, no.1, pp. 163- 171, 1995.

[14] Q. Zhang and S. Kassam, " Finite-state markov model for rayleigh fading channels," *IEEE Transactions on Communications*, vol.47, no.11, pp.1688 -1692, 1999.

[15] L. Lei, C. Lin, J. Cai, and X. Shen, "Flow-level Performance of Opportunistic OFDM-TDMA and OFDMA Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5461-5472, Dec. 2008.

[16] J. Ramis, and G. Femenias, "Cross-Layer Design of Adaptive Multirate Wireless Networks Using Truncated HARQ," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 944 - 954, 2011.

[17] A. Hedayat, and A. Nosratinia, " Outage and Diversity of Linear Receivers in Flat-Fading MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5868 - 5873, 2007.

[18] 3GPP TR 36.814, V9.0.0, "Further advancements for E-UTRA, physical layer aspects," Mar. 2010. [Online]. Available: http://www.3gpp.org

[19] Theodore S. Rappaport, *Wireless Communications: Principles and Practice (2nd Edition)*, Prentice Hall, 2001.

[20] M. Rupp, C. Mecklenbrauker, and G. Gritsch, "High diversity with simple space time block-codes and linear receivers," in Proc. of *IEEE GLOBECOM*, San Francisco, CA, November 2003, pp. 302-306.

[21] B. Rengarajan, C. Caramanis, and G. de Veciana, "Analyzing queueing systems with coupled processors through semidefinite programming," 2008, [Online]. Available: http://users.ece.utexas.edu/ gustavo/papers/ SdpCoupledQs.pdf