

MIMO Broadcast Scheduling with Limited Feedback

Wei Zhang, *Member, IEEE*, and Khaled Ben Letaief, *Fellow, IEEE*

Abstract— We consider multiuser scheduling with limited feedback of partial channel state information in MIMO broadcast channels. By using spatial multiplexing at the base station (BS) and antenna selection for each user, we propose a multiuser scheduling method that allocates independent information streams from all M transmit antennas to the M most favorable users with the highest signal-to-interference-plus-noise ratio (SINR). A close approximation of the achievable sum-rate throughput for the proposed method is obtained and shown to match the simulation results very well. Moreover, two reduced feedback scheduling approaches are proposed. In the first approach, which we shall refer to as selected feedback scheduling, the users are selected based on their SINR compared to a predesigned threshold. Only those selected users are allowed to feed back limited information to the BS. The resultant feedback load and achievable throughput are derived. It will then be demonstrated that with a proper choice of the threshold, the feedback load can be greatly reduced with a negligible performance loss. The second reduced feedback scheduling approach employs quantization for each user, in which only few bits of quantized SINR are fed back to the BS. Performance analysis will show that even with only 1-bit quantization, the proposed quantized feedback scheduling approach can exploit the multiuser diversity at the expense of slight decrease of throughput.

Index Terms— Antenna selection, broadcast channel, feedback, multiple-input multiple-output, multiuser diversity.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) systems with a number of transmit and receive antennas hold the promise of achieving high capacity over wireless channels [1], [2]. One key benefit of MIMO is the resultant *spatial diversity gain* which can be exploited to get a more reliable reception performance. Space-time block coding (STBC) is a MIMO technique which can achieve the maximum spatial diversity gain in an efficient way [3]. Another merit of using MIMO arises from the high *spatial multiplexing gain* which can be obtained by sending independent information streams over any transmit-receive antenna pair simultaneously [4], [5].

In the context of a broadcast channel (BC) in a MIMO cellular system, the spatial diversity gain or multiplexing gain can still be achieved for a single user. From the system point of view, however, the total throughput may not be optimal because a potential inherent *multiuser diversity* is not exploited. Multiuser diversity is a form of *selection diversity*

among users which arises from the independent fading channels between the base station (BS) and the multiple users [6], [7]. To take advantage of the multiuser diversity, a variety of multiuser scheduling algorithms have been proposed for MIMO downlink transmissions in [8]–[14]. But these are mostly based on time-division-multiple-access (TDMA) in which the BS transmits to only one user at a time. It was later shown that the maximum achievable sum-rate using TDMA for MIMO broadcast channels is only a small fraction of the sum-rate capacity of MIMO broadcast channels [15].

Dirty paper coding (DPC) has been shown to be an optimal approach to achieve MIMO broadcast channel capacity by serving multiple users simultaneously [16], [17]. Compared with TDMA, DPC can achieve a linear increase of capacity in $\min(M, K)$ times [15], where M is the number of transmit antennas and K is the number of users. To alleviate the computation burden of DPC, simple sum power iterative water-filling (IWF) algorithms were proposed in [18], [19]. For a large number of users, some practical techniques with reduced complexity to achieve the sum capacity have also been investigated in [20]–[24]. To reduce the amount of channel feedback and the computation complexity, user selections were also considered. Through the use of a simple norm-threshold feedback protocol, the maximal scaling of the sum rate is achievable [25]. In [26], only those users with the singular values above a set threshold are allowed to feed back the channel information to the BS. In [27], a low-complexity suboptimal user selection scheme, which was referred to as the semi-orthogonal user selection (SUS) algorithm, was proposed. When the user selection algorithms in [26] and [27] are combined with zero-forcing (ZF) beamforming, the optimum sum-rate throughput can be asymptotically achieved when K goes to infinity.

The aforementioned DPC and the related low-complexity algorithms require perfect knowledge of the channel state information at the transmitter (CSIT). However, it is not practical to assume that the channel coefficients are perfectly known at the transmitter. Furthermore, with an increasing number of antennas and users, the feedback amount of CSI becomes huge. In situations where the feedback is severely limited, quantization of the information needed at the transmitter seems to be attractive [28]. With finite rate feedback, it was shown in [29] that the transmitter can obtain quantized CSI through a random quantization codebook. Full multiplexing gain can then be realized if the number of feedback bits scales with the signal-to-noise ratio (SNR). In [30], and by assuming perfect knowledge of CSI at the receiver only, the signal-to-interference-plus-noise ratio (SINR) is made available to the

Manuscript received June 1, 2006; revised Jan. 5, 2007. This work was supported in part by the Hong Kong Research Grant Council under grant No. HKUST6250/04E.

W. Zhang and K. B. Letaief are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: eewzhang@ust.hk, eekhaled@ece.ust.hk).

Digital Object Identifier 10.1109/JSAC.2007.070918.

transmitter and only those users with high SINR are allowed to receive information. The throughput is then shown to be asymptotically optimal and scale as $M \log \log(KN)$, where N is the number of receive antennas at each user.

In this paper, we will investigate several SINR-based scheduling algorithms for MIMO broadcast channels in the absence of CSIT. Firstly, we propose a MIMO broadcast scheduling with the feedback of SINR from *all* users, namely *full feedback scheduling*. In this scheme, spatial multiplexing and receive antenna selection are both employed. With antenna selection for each user and feedback of SINR from all users, the BS transmits M independent information streams to M users with high SINR simultaneously. This scheme is similar to the one proposed by Sharif and Hassibi in [30]. But ours differs from [30] in that receive antenna selection is used for each user, whereas in [30] every receive antenna is employed as an *individual* user. Because forcing multiple closely spaced antennas belonging to a single user to receive different signals would require extra cost and power, the antenna selection scheme is preferred. Moreover, it will be shown later that our scheme requires less feedback load without loss of throughput as opposed to [30].

We further propose two MIMO broadcast scheduling algorithms with limited feedback. In the first one, which we shall refer to as *selected feedback scheduling*, users are selected based on their SINR compared to a set threshold. Only those selected users are allowed to send back their SINR to the BS. The achievable sum-rate throughput and the resultant feedback load reduction are then studied over Rayleigh fading channels. Performance analysis will demonstrate that the feedback load can be substantially reduced with a negligible loss of throughput. It should be mentioned that a similar technique, SNR thresholding, was proposed in [31] in the context of single-antenna multiuser scheduling. In our second scheduling algorithm with limited feedback, namely, *quantized feedback scheduling*, only few bits of quantized SINR of each user are sent back to the BS. It should be remarked that quantized feedback was applied to multiuser diversity in single-antenna systems [32]–[34]. In contrast, here the quantization of SINR is employed for each user in a multiuser MIMO system. Performance analysis will show that even with only 1-bit quantization, the achievable sum-rate throughput increases with the increasing number of users.

The rest of this paper is organized as follows. Section II introduces the system model. In Section III, a multiuser scheduling algorithm with SINR feedback from all users is proposed for MIMO broadcast channels and its achievable throughput is obtained numerically. Section IV proposes a selected feedback scheduling algorithm. The achievable sum-rate throughput and the feedback load are derived. Section V proposes a quantized feedback scheduling algorithm. In particular, 1-bit feedback scheduling is investigated. Section VI presents further comparison results of all the proposed scheduling algorithms. Finally, we draw our conclusions in Section VII.

II. SYSTEM MODEL

We consider the downlink of a single-cell wireless system where the BS with M antennas communicates with K

geographically dispersed users, each having N antennas. In a typical cellular system, we often have $K \gg M$ and $M \geq N$. Assume that only J out of K users are allowed to communicate with the BS simultaneously at any given time slot t . Let $\mathcal{A}(t)$ be the set of active users which are assigned a non-zero rate at time slot t . Thus, the cardinality of $\mathcal{A}(t)$ is $|\mathcal{A}(t)| = J$ with $1 \leq J \leq K$. Let user k ($1 \leq k \leq K$) be an active user at time slot t , whose received signal is given by

$$\mathbf{Y}_k^t = \sqrt{\alpha_k} \mathbf{H}_k^t \mathbf{X}^t + \mathbf{W}_k^t, \quad k \in \mathcal{A}(t), \quad (1)$$

where \mathbf{Y}_k^t is an $N \times 1$ vector of the received signal at the k th user in time slot t , \mathbf{X}^t is an $M \times 1$ vector of the transmitted signal in time slot t , and \mathbf{W}_k^t is an $N \times 1$ vector of additive noise whose entries are i.i.d. complex Gaussian with zero mean and variance N_0 . \mathbf{H}_k^t is an $N \times M$ channel matrix, where the (n, m) th entry $h_k^t(n, m) \sim \mathcal{CN}(0, 1)$ is i.i.d. complex Gaussian with zero mean and unit variance and represents the complex channel gain from the transmit antenna m to the receive antenna n of user k at time slot t . The scalar α_k models the power attenuation due to the path loss and shadow fading.

We assume that the total transmit power is equal to 1, i.e., $\text{Tr}(\mathbf{X}^t [\mathbf{X}^t]^H) = 1$, where $\text{Tr}(\cdot)$ stands for the trace of a matrix. This ensures that the total transmit power is not dependent on the number of transmit antennas. The block-fading channel model is considered in which the channel coefficient $h_k^t(n, m)$, $\forall m, n, k$ is quasi-static and frequency non-selective in time slot t but varies independently in different time slots. Furthermore, the channel $h_k^t(n, m)$ is uncorrelated with other channels $h_{k'}^t(n', m')$ for any two distinct pairs (m', n', k') and (m, n, k) . Throughout this paper, we assume that the MIMO channels are perfectly known at the receiver, but are unknown at the transmitter.

We may regard model (1) as a general description of several scenarios in downlink transmissions as follows:

- If $J = K = 1$, (1) represents a single user MIMO Gaussian channel [1];
- If $J = K > 1$, (1) represents a MIMO Gaussian broadcast channel [16];
- If $J = 1$, and $K > 1$, (1) represents a *multiuser scheduling* MIMO Gaussian channel, where only one user is scheduled at a given time [10]–[13];
- If $1 < J < K$, (1) represents a *multiuser scheduling* MIMO Gaussian channel, where more than one user are scheduled at a given time [20], [26], [27], [30], [35].

In this paper, we focus on the problem of *multiuser scheduling* in the last scenario, where multiple users are scheduled simultaneously at any time.

The symbol $x(m)$ is transmitted through antenna m with $x(m)$ assumed to be independent for $m = 1, \dots, M$. The average transmit power per antenna is $\frac{1}{M}$, i.e. $E[|x(m)|^2] = \frac{1}{M}$, thereby ensuring that the total transmit power is 1. The received signal at the n th antenna of user k can then be derived from (1) as

$$y_k(n) = \sqrt{\alpha_k} \sum_{m=1}^M h_k(n, m) x(m) + w_k(n). \quad (2)$$

Assume that the signal $x(m)$ is the desired signal for user k . By regarding other signals $x(m')$, $m' \neq m$ as interference, we

can get the *instantaneous* SINR of $y_k(n)$ as

$$\text{SINR}_{m,n}^{(k)} \triangleq \frac{|h_k(n, m)|^2}{\frac{M}{\bar{\gamma}_k} + \sum_{m' \neq m} |h_k(n, m')|^2}, \quad (3)$$

where $\bar{\gamma}_k = \alpha_k/N_0$ is the average SNR of user k .

In [30], a downlink scheduling scheme is proposed with SINR feedback. For the MIMO case, it treats each receive antenna at any receiver as an individual user, so that there are effectively NK single antenna receivers. With the feedback of the maximum SINR of each user along with the index m which achieves the maximum SINR, the BS will transmit M independent signals to M users with the highest SINR simultaneously. Let $\mathcal{S}_t = \{1, \dots, M\}$, $\mathcal{S}_r = \{1, \dots, N\}$ and $\mathcal{S}_u = \{1, \dots, K\}$. The achievable throughput R can then be bounded by [30]

$$R \leq E \left[\sum_{m=1}^M \log_2 \left(1 + \max_{n \in \mathcal{S}_r, k \in \mathcal{S}_u} \text{SINR}_{m,n}^{(k)} \right) \right]. \quad (4)$$

The asymptotic analysis in [30] shows that the achievable throughput is $M \log \log(NK)$ when K goes to infinity. However, the SINR-based scheduling algorithm in [30] requires the feedback of KN SINR values and the feedback load increases with the increase of the number of receive antennas.

III. MIMO BROADCAST SCHEDULING WITH SINR FEEDBACK

In this section, we propose a MIMO downlink scheduling algorithm with antenna selection. It only requires a feedback of total K SINR values. Performance analysis shows that the reduced feedback scheduling does not induce any loss of throughput and fairness compared with [30].

A. Scheduling Algorithm

The proposed scheduling algorithm with the full feedback of SINR can be described as a two-stage algorithm: *feedback* by users and *scheduling* by the BS.

Stage 1: Feedback

- 1) Initialization: Set $B_k = 0$, $\mathbf{T}\mathbf{x}_k = 0$ and $\mathbf{R}\mathbf{x}_k = 0$ for $k = 1, \dots, K$.
- 2) For $k = 1$ to K ,

$$B_k = \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$

$$\mathbf{T}\mathbf{x}_k = \arg_m \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$

$$\mathbf{R}\mathbf{x}_k = \arg_n \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$
 User k feeds back to the BS $\mathbf{T}\mathbf{x}_k$ and B_k .

End For

Stage 2: Scheduling

- 1) Initialization: Set $k_m^* = 0$ and $\mathcal{I}_m = \emptyset$, for $m = 1, \dots, M$. Get the feedback of $\mathbf{T}\mathbf{x}_k$ and B_k .
- 2) For $k = 1$ to K ,

$$m = \mathbf{T}\mathbf{x}_k,$$

$$\mathcal{I}_m = \mathcal{I}_m \cup k.$$
 End For
- 3) For $m = 1$ to M ,
 If $\mathcal{I}_m \neq \emptyset$

$$k_m^* = \arg \max_{k \in \mathcal{I}_m} B_k.$$
 End If

End For

The key idea behind the above scheduling algorithm is that each user firstly selects the most favorable transmit antenna $\mathbf{T}\mathbf{x}_k$ and receive antenna $\mathbf{R}\mathbf{x}_k$ for achieving the maximum SINR B_k and then feeds the corresponding B_k along with the index $\mathbf{T}\mathbf{x}_k$ to the BS. Since receive antenna selection is performed at each user, the total feedback load at each time slot is K real values $\{B_1, \dots, B_K\}$ plus $K \lceil \log_2 M \rceil$ bits of $\mathbf{T}\mathbf{x}_k, \forall k$, where $\lceil x \rceil$ denotes the smallest integer larger than x . At the BS, the users with $\mathbf{T}\mathbf{x}_k = m$ are clustered into the set \mathcal{I}_m and then the scheduled user k_m^* is decided by achieving $\max_{k \in \mathcal{I}_m} B_k$. Once users k_m^* are determined for all $m = 1, \dots, M$, the receive antenna $\mathbf{R}\mathbf{x}_{k_m^*}$ of user k_m^* is selected to receive the signal $x(m)$.

B. Throughput Analysis

In the following, we will derive the average achievable throughput of the proposed scheduling algorithm over Rayleigh fading channels.

Under Gaussian codes and the minimum distance decoding at receiver, the instantaneous sum-rate of the scheduling with the feedback of SINR is

$$R = \sum_{m=1}^M R_m = \sum_{m=1}^M \log_2 \left(1 + \text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)} \right), \quad (5)$$

where $\text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)}$ is the SINR of user k_m^* scheduled by the m th transmit antenna, and $\mathbf{R}\mathbf{x}_{k_m^*}$ is the receive antenna selected by user k_m^* . Note that

$$\begin{aligned} \text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)} &= \max_{k \in \mathcal{I}_m} B_k \\ &= \max_{k \in \mathcal{I}_m} \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)} \\ &\approx \max_{l \in \{1, \dots, NK\}} Z_l, \end{aligned} \quad (6)$$

where

$$Z_l \triangleq \frac{|g_l|^2}{\frac{M}{\bar{\gamma}_l} + \sum_{j=1}^{M-1} |\tilde{g}_{l,j}|^2} \quad (7)$$

with $g_l, \tilde{g}_{l,j} \sim \mathcal{CN}(0, 1)$ for all $l = 1, \dots, NK$ and $j = 1, \dots, M-1$. Considering that $\mathcal{I}_m = \{k | \mathbf{T}\mathbf{x}_k = m, \forall k \in \{1, \dots, K\}\}$ and $\text{Prob}\{\mathbf{T}\mathbf{x}_k = m\} = \frac{1}{M}$ for any k and m , we can get the average of cardinality $E|\mathcal{I}_m| \approx K/M$. Then, the SINR of the scheduled user is selected from $E|\mathcal{I}_m| \cdot |\mathcal{S}_t| \cdot |\mathcal{S}_r| \approx NK$ candidates as shown in the last step of (6).

Since Z_l 's are i.i.d, we can rewrite (7) as $Z = \frac{X}{\frac{M}{\bar{\gamma}} + Y}$, where $X \sim \chi^2(2)$ and $Y \sim \chi^2(2M-2)$. Then, the PDF and the CDF of Z_l can be given by [30]

$$f_{Z_l}(t) = \frac{e^{-Mt/\bar{\gamma}_l}}{(1+t)^M} \left(\frac{M}{\bar{\gamma}_l} (1+t) + M-1 \right), \quad t \geq 0 \quad (8)$$

$$F_{Z_l}(t) = 1 - \frac{e^{-Mt/\bar{\gamma}_l}}{(1+t)^{M-1}}, \quad t \geq 0. \quad (9)$$

Let

$$U = \max_{l \in \{1, \dots, NK\}} Z_l \quad (10)$$

$$P = \max_{k \in \mathcal{I}_m} \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)}. \quad (11)$$

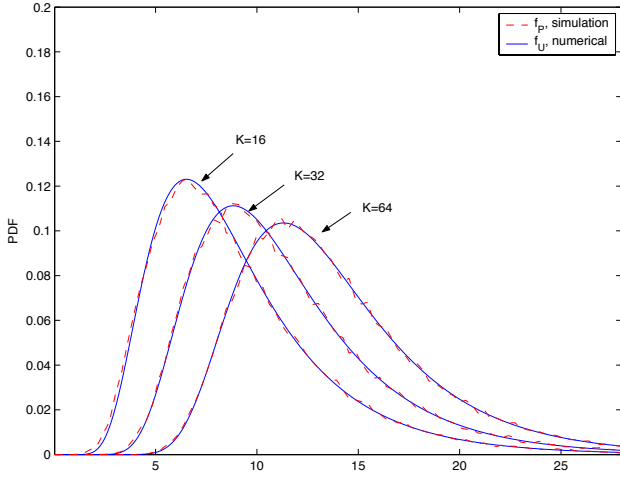


Fig. 1. PDF comparison between P and U for $K = 16, 32$ and 64 at 10 dB SNR for $M = 2$ and $N = 2$. Dashed curves denote the simulation results of the PDF of P . Solid curves denote the numerical results of the PDF of U .

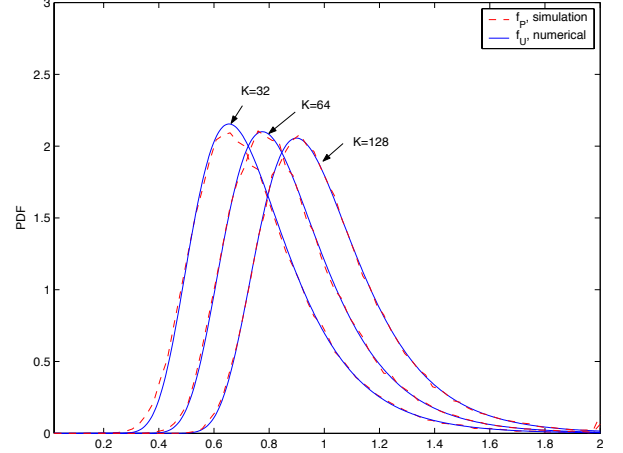


Fig. 2. PDF comparison between P and U for $K = 32, 64$ and 128 at 0 dB SNR for $M = 4$ and $N = 2$. Dashed curves denote the simulation results of the PDF of P . Solid curves denote the numerical results of the PDF of U .

Then, the CDF of U can be easily given by [36]

$$F_U(t) = \prod_{l=1}^{NK} F_{Z_l}(t).$$

The PDF of U is given by

$$f_U(t) = \frac{d}{dt} F_U(t) = \sum_{l=1}^{NK} \left(f_{Z_l}(t) \prod_{j=1, \dots, NK; j \neq l} F_{Z_j}(t) \right). \quad (12)$$

Using (5) and (6), the system throughput averaged over Rayleigh fading channels is given by

$$\begin{aligned} E(R) &= \sum_{m=1}^M E \log_2 \left(1 + \text{SINR}_{m, \text{Rx}_{k_m}^*}^{(k_m^*)} \right) \\ &= M \int_0^\infty \log_2(1+t) f_P(t) dt \\ &\approx M \int_0^\infty \log_2(1+t) f_U(t) dt. \end{aligned} \quad (13)$$

In heterogeneous networks, the achievable throughput of the proposed scheduling algorithm can be obtained by substituting (12) into (13). Consequently, we get

$$\begin{aligned} E(R) &\approx M \sum_{l=1}^{NK} \int_0^\infty \log_2(1+t) f_{Z_l}(t) \\ &\quad \cdot \prod_{j=1, \dots, NK; j \neq l} F_{Z_j}(t) dt \end{aligned} \quad (14)$$

where $f_{Z_l}(t)$ and $F_{Z_j}(t)$ are given by (8) and (9), respectively.

In homogeneous networks, all users have an identical $\bar{\gamma}$. Then, (12) reduces to

$$f_U(t) = NK f_Z(t) (F_Z(t))^{NK-1}. \quad (15)$$

Hence, the achievable throughput of the MIMO downlink scheduling averaged over fading channels is

$$E(R) \approx KMN \int_0^\infty \log_2(1+t) f_Z(t) (F_Z(t))^{NK-1} dt. \quad (16)$$

The numerical approximation of the throughput given in (16) results from the approximation $P \approx U$ in (6). Since the

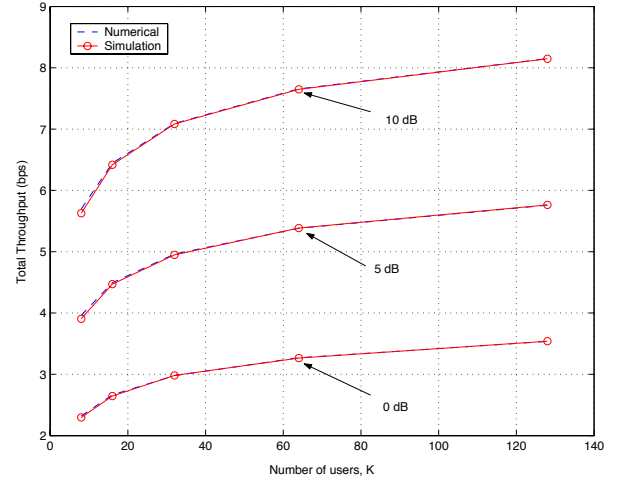


Fig. 3. Total throughput of the full feedback multiuser scheduling vs. the number of users at various SNR, 0, 5 and 10 dB for $M = 2$ and $N = 2$.

PDF of U is used to approximate the PDF of P in calculating the throughput in (13), we should evaluate the accuracy of the approximation.

In Fig. 1, we compare the PDF $f_U(t)$ and $f_P(t)$ at 10 dB SNR with $M = 2$ and $N = 2$ for 16, 32 and 64 users, respectively. The PDF $f_U(t)$ is obtained numerically from (15), whereas the PDF $f_P(t)$ is plotted by simulating P with 100,000 random runs according to (11). It can be seen from Fig. 1 that the PDF $f_U(t)$ matches the PDF $f_P(t)$ very well. Thus, the approximation in (13) is very tight. Similar observations can be obtained from Fig. 2 which shows the PDF $f_U(t)$ and $f_P(t)$ at 0 dB SNR with $M = 4$ and $N = 2$ for 32, 64 and 128 users, respectively.

We further perform the simulation of the MIMO downlink scheduling algorithm with $M = 2, N = 2$ for 0, 5 and 10 dB SNR, respectively. The simulation results of the achievable throughput are compared with the numerical results and are shown in Fig. 3. Such figure demonstrates that the numerical results obtained from (16) are accurate enough to characterize

the simulation results for all examined SNR values and K .

C. Discussion

In [30], three signal reception approaches were proposed for the case of multiple antennas at each receiver. It was then shown to achieve the largest throughput, each receive antenna should be treated as an individual user. By doing so, the K -user $M \times N$ system is converted to (NK) -user $M \times 1$ system [30]. Note in the case of single antenna for each user, our approach is the same as the one in [30] because no receive antenna selection is needed. Thus, (16) can also be used to evaluate the performance of the approach in [30] for (NK) -user $M \times 1$ system. Then, we get

$$E(R) \approx NKM \int_0^\infty \log_2(1+t) f_Z(t) (F_Z(t))^{NK-1} dt. \quad (17)$$

It should be mentioned that (16) is the achievable throughput of our approach using receive antenna selection at each user, whereas (17) is the achievable throughput of the approach in [30] which treats each receive antenna as an individual user. After observing that the right hand side (RHS) of (16) is identical to the RHS of (17), we can see that our approach does not induce any performance loss compared with the best approach proposed in [30].

Next, we will show that our algorithm is in fact equivalent to the algorithm of [30] in terms of performance though they are different in terms of receiver design. As shown in (5), for the signal transmitted from the m th antenna, the best user k^* is selected with the largest $\text{SINR}_{m, \mathbf{R}_{k^*}}^{(k^*)}$. Furthermore, from (6) we see that $\text{SINR}_{m, \mathbf{R}_{k^*}}^{(k^*)}$ can be approximated by $\max_{l \in \{1, \dots, NK\}} Z_l$, i.e., $P \approx U$ where U and P are given in (10) and (11), respectively. The close approximation of the PDFs of U and P along with the resultant throughput has been validated. Thus, we can conclude that our scheduling algorithm with antenna selection is equivalent to the algorithm of [30] that chooses the user with the largest SINR from all NK users each with single antenna.

In heterogeneous networks, the users usually have various SNRs due to their different distances from the base station. A scheduling algorithm that only takes advantage of the best channel gain will then induce the unfairness of rate allocation in heterogeneous networks, even though it can maximize the system throughput. This is because the user located far from the BS will be seldom scheduled and the user near to the BS will dominate the resources. Therefore, the fairness issue is a major concern for the design of practical multiuser MIMO systems. Next, we would like to evaluate the fairness of our proposed scheduling algorithm based on the SINR feedback with antenna selection.

Assume that user k^* is the weakest user with the smallest SINR, i.e., $\bar{\gamma}_{k^*} = \bar{\gamma}_{\min} = \min_{k \in \{1, 2, \dots, K\}} \bar{\gamma}_k$. Let $P_{\bar{\gamma}_{\min}, m, \mathbf{R}_{k^*}}$ be the probability of the event that antenna \mathbf{R}_{k^*} of the weakest user k^* is scheduled for the signal transmitted from the m th transmit antenna, which is given by

$$P_{\bar{\gamma}_{\min}, m, \mathbf{R}_{k^*}} = \text{Prob}\{\text{SINR}_{m, \mathbf{R}_{k^*}}^{(k^*)} = \max_{k \in \mathcal{I}_m} \max_{n \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m, n}^{(k)}\} \quad (18)$$

Using (6), we further obtain

$$P_{\bar{\gamma}_{\min}, m, \mathbf{R}_{k^*}} \approx \text{Prob}\{Z_j > Z_1, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_{NK}\},$$

where Z_j is the SINR of the virtual user j in the NK -user $M \times 1$ system corresponding to the antenna \mathbf{R}_{k^*} of the weakest user k^* in the K -user $M \times N$ system after treating all NK receive antennas as individual users. Z_l is given by (7) and its PDF is given by (8). Next, we recall the main result of [30, Theorem 6] and get

$$P_{\bar{\gamma}_{\min}, m, \mathbf{R}_{k^*}} \geq \frac{1}{NK} \frac{M-1 + \frac{M}{\bar{\gamma}_{\min}}}{M-1 + \frac{M}{\bar{\gamma}_{\max}}} \cdot e^{-\left(\frac{M}{\bar{\gamma}_{\min}} - \frac{M}{\bar{\gamma}_{\max}}\right) \left(e^{\frac{2 \log(NK)}{M-1}} - 1\right)} + O(1/(NK)^2),$$

where $\bar{\gamma}_{\max} = \max_{k \in \{1, 2, \dots, K\}} \bar{\gamma}_k$. When $\frac{\log(NK)}{M}$ becomes very small, $P_{\bar{\gamma}_{\min}, m, \mathbf{R}_{k^*}} \rightarrow \frac{1}{NK}$. This implies that the antenna \mathbf{R}_{k^*} of the weakest user has a probability of $\frac{1}{NK}$ to be scheduled by the signal transmitted from the m th antenna. Note that $\text{SINR}_{m, n}^{(k^*)}$ for $m = 1, \dots, M$ and $n = 1, \dots, N$ have identical distributions, $P_{\bar{\gamma}_{\min}, m, n}$, which do not depend on the indices of the transmit antenna m and receive antenna n . Therefore, each antenna of the weakest user has the same probability to access the system resources. In other words, the weakest user has a probability of $\frac{1}{K}$ to be scheduled by the proposed scheduling algorithm with antenna selection. It achieves the same fairness as that of [30].

Finally, we should stress that our scheduling with antenna selection does not induce any loss of throughput nor fairness compared with [30]. However, from the point of view of practical implementation, receive antenna selection for MIMO downlink scheduling proposed in this paper is much more favorable.

IV. MIMO BROADCAST SCHEDULING WITH SELECTED FEEDBACK

In this section, a scheduling algorithm with selected feedback of SINR is proposed for MIMO broadcast channels. The achievable throughput and the feedback load are also analyzed.

A. Scheduling Algorithm

The proposed scheduling algorithm with the selected feedback of SINR can be described as a two-stage algorithm: *selected feedback* by users and *scheduling* by the BS.

Stage 1: Selected Feedback

- 1) Initialization: Set $B_k = 0$, $\text{Tx}_k = 0$ and $\mathbf{R}_k = 0$ for $k = 1, \dots, K$.
- 2) For $k = 1$ to K ,
 - $B_k = \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m, n}^{(k)}$,
 - $\text{Tx}_k = \arg_m \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m, n}^{(k)}$,
 - $\mathbf{R}_k = \arg_n \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m, n}^{(k)}$,
 - If $B_k \geq \lambda$,
 - user k is selected to feed back to the BS Tx_k and B_k .
 - Else,
 - user k does not feed back anything.

End If
 End For
 Stage 2: Scheduling
 1) Initialization: Set $k_m^* = 0$ and $\mathcal{I}_m = \emptyset$, for $m = 1, \dots, M$. Get the feedback of B_k and the corresponding $\mathbf{T}\mathbf{x}_k$ from the selected user k which has $B_k \geq \lambda$.
 2) For every selected user k ,
 $m = \mathbf{T}\mathbf{x}_k$,
 $\mathcal{I}_m = \mathcal{I}_m \cup k$.
 End For
 3) For $m = 1$ to M ,
 If $\mathcal{I}_m \neq \emptyset$
 $k_m^* = \arg \max_{k \in \mathcal{I}_m} B_k$.
 End If
 End For

The principal idea behind the selected feedback scheduling is that only those users whose maximum SINR are larger than a predesigned threshold λ can be selected to feed back the SINR to the BS. The advantage of this algorithm is that a low feedback is obtained due to the user selection.

B. Throughput Analysis

Note that \mathcal{I}_m is the set of *selected* users preferring to the m th transmit antenna. Let \mathcal{S}_m denote the set of *all* users having $\mathbf{T}\mathbf{x}_k = m$, i.e., $\mathcal{S}_m = \{k | \mathbf{T}\mathbf{x}_k = m, \forall k \in \{1, \dots, K\}\}$. Obviously, $\mathcal{I}_m \subseteq \mathcal{S}_m$ because $\mathcal{I}_m = \{k | B_k \geq \lambda, \forall k \in \mathcal{S}_m\}$. The system throughput averaged over Rayleigh fading channels is then given by

$$E(R) = \sum_{m=1}^M E \log_2 \left(1 + \text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)} \right), \quad (19)$$

where

$$\begin{aligned} \text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)} &= \max_{k \in \mathcal{I}_m} B_k \\ &= \max_{k \in \mathcal{S}_m} B_k \\ &\approx \max_{l \in \{1, \dots, NK\}} Z_l. \end{aligned} \quad (20)$$

By substituting (20) into (19) and using (10), we can obtain

$$E(R) \approx M \int_{\lambda}^{\infty} \log_2(1+t) f_U(t) dt. \quad (21)$$

After substituting (12) into (21), we can get the achievable throughput in heterogeneous networks as

$$E(R) \approx M \sum_{l=1}^{NK} \int_{\lambda}^{\infty} \log_2(1+t) f_{Z_l}(t) \prod_{j=1, \dots, NK; j \neq l} F_{Z_j}(t) dt$$

where $f_{Z_l}(t)$ and $F_{Z_j}(t)$ are given by (8) and (9), respectively. In homogeneous networks, we can substitute (15) into (21) and obtain

$$E(R) \approx KMN \int_{\lambda}^{\infty} \log_2(1+t) f_Z(t) (F_Z(t))^{NK-1} dt. \quad (22)$$

It can be observed that when $\lambda \rightarrow 0$, (22) is equivalent to (16).

To investigate the accuracy of the approximation in (22), we perform simulation of the selected feedback scheduling

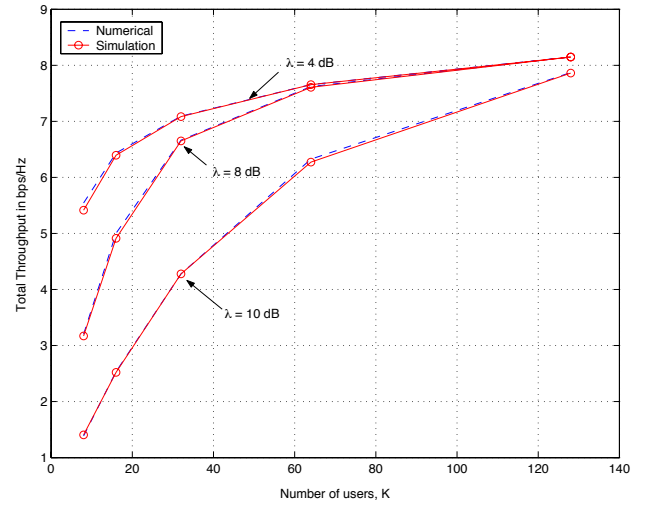


Fig. 4. Total throughput of the selected feedback multiuser scheduling vs. the number of users for various threshold $\lambda = 4, 8$ and 10 dB. $M = 2$ and $N = 2$. SNR is 10 dB.

with $M = N = 2$ at 10 dB SNR for $\lambda = 4, 8$ and 10 dB, respectively. The simulation results are compared with the numerical results achieved by (22) and are shown in Fig. 4. We can see that the numerical results match the simulation results with negligible difference. Furthermore, with the increase of the threshold, the throughput is decreased. This is reasonable because less users are allowed to participate in the feedback when the threshold is increased.

C. Feedback Load Analysis

Assume that l users are selected for feedback in one time slot. This indicates that there are l users satisfying $B_k \geq \lambda$. Since $B_k = \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)} \forall k$ are i.i.d., we can denote $F_B(t)$ the CDF of B_k . Then, the probability of selecting l users in the current time slot is

$$\mathcal{P}_l = \binom{K}{l} (1 - F_B(\lambda))^l (F_B(\lambda))^{K-l} \quad (23)$$

With the notation of Z_l in (7), we can obtain

$$B_k \approx \max_{l \in \{1, \dots, MN\}} Z_l \quad (24)$$

and

$$F_B(\lambda) \approx [F_Z(\lambda)]^{MN}. \quad (25)$$

The average feedback load of the selected scheduling is given by

$$\mathcal{L} = \sum_{l=1}^K l \cdot \mathcal{P}_l. \quad (26)$$

To compare with the full feedback scheduling, we define the average feedback load ratio (FLR) ζ as

$$\zeta \triangleq \frac{\mathcal{L}}{K}. \quad (27)$$

Using (23)-(26), we can rewrite (27) as

$$\zeta \approx \sum_{l=1}^K \frac{l}{K} \binom{K}{l} (1 - [F_Z(\lambda)]^{MN})^l ([F_Z(\lambda)]^{MN})^{K-l}, \quad (28)$$

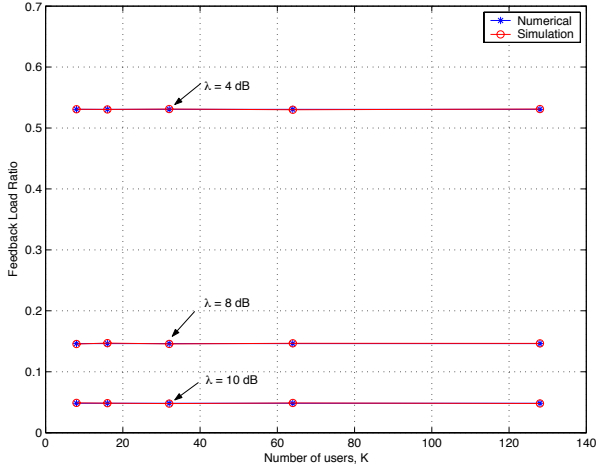


Fig. 5. Feedback load ratio of the selected feedback multiuser scheduling vs. the number of users for various threshold $\lambda = 4, 8$ and 10 dB. $M = 2$ and $N = 2$. SNR is 10 dB.

where $F_Z(\lambda)$ is given by (9).

Note that using

$$\frac{l}{K} \binom{K}{l} = \binom{K-1}{l-1}, \quad (29)$$

we can rewrite (28) as

$$\begin{aligned} \zeta &\approx \sum_{l=1}^K \binom{K-1}{l-1} (1 - [F_Z(\lambda)]^{MN})^l ([F_Z(\lambda)]^{MN})^{K-l} \\ &= \sum_{l'=0}^{K-1} \binom{K-1}{l'} (1 - [F_Z(\lambda)]^{MN})^{l'} \\ &\quad ([F_Z(\lambda)]^{MN})^{(K-1)-l'} (1 - [F_Z(\lambda)]^{MN}) \\ &= 1 - [F_Z(\lambda)]^{MN}. \end{aligned} \quad (30)$$

From (30), we can see that the FLR is not dependent on the number of users K .

Fig. 5 shows the comparison in the average FLR for $\lambda = 4, 8$ and 10 dB at 10 dB SNR, respectively. Those results demonstrate that the numerical approximation given by (30) is very accurate. Furthermore, it can be observed that the FLR is not dependent on the number of users. Also, when the threshold is increased, FLR decreases.

D. Throughput-FLR Tradeoff

From (22) and (30), it can be seen that the throughput and FLR both depend on the threshold λ and decrease when λ increases. Therefore, a tradeoff exists between increasing the throughput and decreasing FLR. In practice, two schemes may be used: *throughput-oriented* and *FLR-oriented*. The former scheme is to minimize FLR while guaranteeing a target throughput. The latter scheme is to maximize the throughput while attaining a target FLR. Since the numerical solutions (22) and (30) closely approximate the throughput and FLR, respectively, they can be efficiently employed to investigate the throughput-FLR tradeoff performance.

Fig. 6 shows the throughput-threshold curve using (22) for the *throughput-oriented* scheme. On the other hand, Fig.

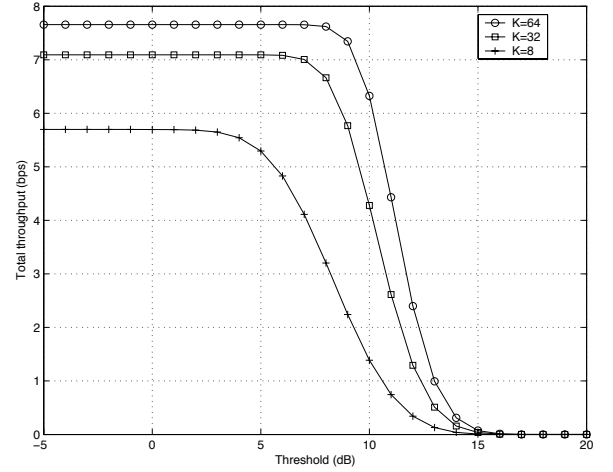


Fig. 6. Total throughput of the selected feedback multiuser scheduling vs. threshold λ (dB) for various number of users $K = 8, 32$ and 64 . $M = 2$ and $N = 2$. SNR is 10 dB.

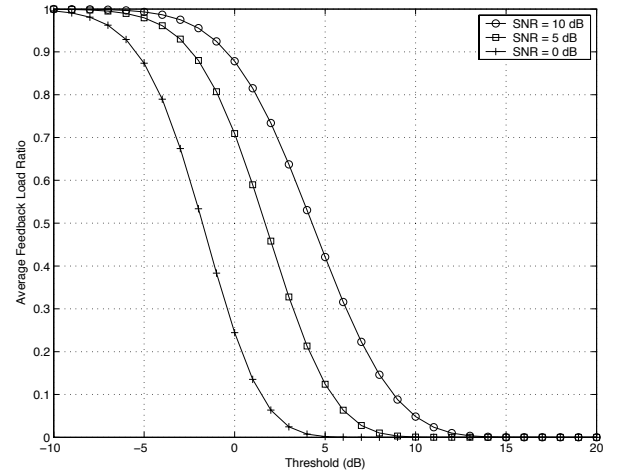


Fig. 7. Feedback load ratio of the selected feedback multiuser scheduling vs. threshold λ (dB) at various SNR $\bar{\gamma} = 0, 5$ and 10 dB. $M = 2$ and $N = 2$.

7 shows the FLR-threshold curve using (30) for the *FLR-oriented* scheme. Given a target throughput, we could find the maximum threshold from Fig. 6 and then use the threshold to get the corresponding FLR from Fig. 7 which is the minimum FLR to get the target throughput. For instance, in the case of $M = N = 2$ with $K = 64$ at 10 dB SNR, the target throughput is 6.3 bps. Firstly, we can read from Fig. 6 that the threshold corresponding to the target throughput is 10 dB. Then, from Fig. 7, we get that the corresponding FLR is 0.05 . Given a target FLR, say 0.4 for $M = N = 2$ at 10 dB SNR, we can observe from Fig. 7 that the corresponding threshold is about 5 dB which is not dependent on K . From Fig. 6, we can also see that the throughput is 7.7 bps which is also the maximum achievable throughput. Therefore, we can conclude that FLR can be greatly reduced without sacrificing the throughput. Finally, in Fig. 8, we show the system throughput with respect to the required FLR for various number of users. It shows that the FLR can be reduced to 0.2 without loss of system throughput for more than 32 users when

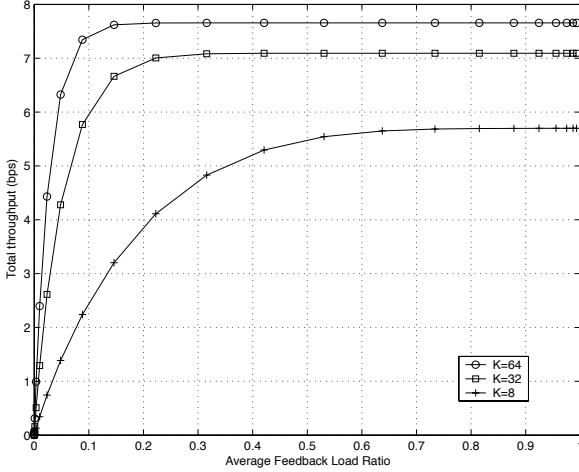


Fig. 8. Tradeoff between throughput and FLR of the selected feedback multiuser scheduling for various number of users at SNR $\bar{\gamma} = 10$ dB. $M = 2$ and $N = 2$.

$M = N = 2$ at 10 dB SNR.

V. MIMO BROADCAST SCHEDULING WITH QUANTIZED FEEDBACK

In this section, a scheduling algorithm with quantized feedback of SINR is proposed for MIMO broadcast channels. The achievable throughput is also derived and in particular scheduling with 1-bit feedback is investigated.

A. Scheduling Algorithm

The proposed scheduling algorithm with the quantized feedback of SINR can be described as a two-stage algorithm: *quantized feedback* by users and *scheduling* by the BS.

Stage 1: Quantized Feedback

- 1) Initialization: Set $B_k = 0$, $\mathbf{T}\mathbf{x}_k = 0$, $\mathbf{R}\mathbf{x}_k = 0$ and $q_k = 0$ for $k = 1, \dots, K$.
- 2) For $k = 1$ to K ,

$$B_k = \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$

$$\mathbf{T}\mathbf{x}_k = \arg_m \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$

$$\mathbf{R}\mathbf{x}_k = \arg_n \max_{m \in \mathcal{S}_t, n \in \mathcal{S}_r} \text{SINR}_{m,n}^{(k)},$$

$$q_k = Q(B_k) \text{ with } q_k \in \{0, 1, \dots, 2^b - 1\},$$
 User k feeds back to the BS $\mathbf{T}\mathbf{x}_k$ and q_k .

End For

Stage 2: Scheduling

- 1) Initialization: Set $k_m^* = 0$ and $\mathcal{I}_m, \mathcal{J}_m = \emptyset$, for $m = 1, \dots, M$. Get the feedback of $\mathbf{T}\mathbf{x}_k$ and q_k from the user k , $k = 1, 2, \dots, K$.
 - 2) For $k = 1$ to K ,

$$m = \mathbf{T}\mathbf{x}_k,$$

$$\mathcal{I}_m = \mathcal{I}_m \cup k.$$
 End For
 - 3) For $m = 1$ to M ,
 If $\mathcal{I}_m \neq \emptyset$

$$\mathcal{J}_m = \{\tilde{k} | \tilde{k} = \arg_k \max_{k \in \mathcal{I}_m} q_k\},$$

$$k_m^* = \text{rand}\{\mathcal{J}_m\}.$$
 End If
- End For

Compared with the full feedback scheduling where each user feeds a real value B_k to the BS, the quantized feedback scheduling requires each user to send back a quantized value $Q(B_k)$ which can be represented by few bits. In fact, a real value B_k is quantized into one of a few levels $q_k = Q(B_k)$ where

$$Q(B_k) = \begin{cases} 0, & 0 \leq B_k < \lambda_1 \\ i, & \lambda_i \leq B_k < \lambda_{i+1}, \quad i = 1, \dots, L-2 \\ L-1, & B_k \geq \lambda_{L-1} \end{cases}$$

The number of levels L is determined by the number of bits b required to represent a value B_k and $L = 2^b$.

With the feedback of q_k and $\mathbf{T}\mathbf{x}_k$ from all users, the BS will firstly cluster the users having $\mathbf{T}\mathbf{x}_k = m$ into the set \mathcal{I}_m and then cluster the users achieving the maximum of q_k for all $k \in \mathcal{I}_m$ into the set \mathcal{J}_m , i.e. $\mathcal{J}_m = \{\tilde{k} | \tilde{k} = \arg_k \max_{k \in \mathcal{I}_m} q_k\}$. Obviously, $\mathcal{J}_m \subseteq \mathcal{I}_m$. Finally, the transmit antenna m is allocated to user k_m^* which is randomly selected from the set \mathcal{J}_m .

B. Throughput Analysis

The instantaneous sum-rate achieved by the quantized feedback scheduling is given by

$$\begin{aligned} R &= \sum_{m=1}^M \log_2 \left(1 + \text{SINR}_{m, \mathbf{R}\mathbf{x}_{k_m^*}}^{(k_m^*)} \right) \\ &= \sum_{m=1}^M \log_2 (1 + \text{rand}\{B_k, \forall k \in \mathcal{J}_m\}) \end{aligned} \quad (31)$$

where $\text{rand}\{B_k, \forall k \in \mathcal{J}_m\}$ denotes the random selection of an element from the set $\{B_k, \forall k \in \mathcal{J}_m\}$. Note that B_k is approximated by (24) and the CDF of B is given by (25).

Let $V = \text{rand}\{B_k, \forall k \in \mathcal{J}_m\}$. The average total throughput is then given by

$$E(R) = M \int_0^\infty \log_2(1+v) f_V(v) dv, \quad (32)$$

where $f_V(v)$ denotes the PDF of V and will be derived next.

Let $F_V(v)$ denote the CDF of V and $K' = \frac{K}{M}$. Note that $\text{SINR}_{m,n}^{(k)}$ in (3) for $m = 1, 2, \dots, M$ have the same distribution, with each user having an identical chance to select different transmit antennas, i.e., $\text{Prob}\{\mathbf{T}\mathbf{x}_k = m\} = \frac{1}{M}$. Then, for a large number of users we can get that the cardinality of the set \mathcal{I}_m is equal to K' for all $m = 1, 2, \dots, M$. Let A_i denote the range $[\lambda_i, \lambda_{i+1})$ for $i = 0, 1, \dots, L-1$, where $\lambda_0 = 0$ and $\lambda_L = \infty$. Then, we can derive the CDF F_V in the following two cases:

- When $0 \leq V < \lambda_1$, B_k of the scheduled user is in range A_0 . In this case, B_k for all $k \in \mathcal{I}_m$ are also in range A_0 . Thus, the CDF of V is

$$F_V(v) = F_B(v) [F_B(\lambda_1)]^{K'-1}, \quad 0 \leq v < \lambda_1$$

where $F_B(v)$ denotes the CDF of B and is given by (25).

- When $\lambda_i \leq V < \lambda_{i+1}$ for $i = 1, \dots, L-1$, B_k of the scheduled user is in range A_i . We can assume that there are other $(r-1)$ users with B_k in region A_i for any

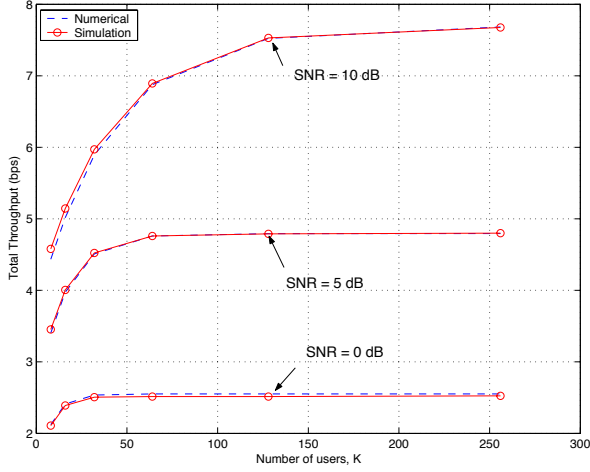


Fig. 9. Total throughput of the quantized feedback multiuser scheduling (1-bit quantization) vs. the number of users at various SNR, 0, 5 and 10 dB. Quantization threshold is fixed to be equal to the SNR. $M = 2$ and $N = 2$.

$1 \leq r \leq K'$. For the remaining $(K' - r)$ users, their B_k must be smaller than λ_i . Thus, the CDF of V is

$$F_V(v) = [F_B(\lambda_i)]^{K'} + \sum_{r=1}^{K'} \binom{K'}{r} [F_B(v) - F_B(\lambda_i)]^{r-1} \cdot [F_B(\lambda_i)]^{K'-r} [F_B(\lambda_{i+1}) - F_B(\lambda_i)]^{r-1},$$

for $\lambda_i \leq v < \lambda_{i+1}$.

Consequently, the PDF $f_V(v)$ is given by

$$f_V(v) = \begin{cases} f_B(v)[F_B(\lambda_1)]^{K'-1}, & \text{for } 0 \leq v < \lambda_1 \\ \sum_{r=1}^{K'} \binom{K'}{r} [F_B(\lambda_i)]^{K'-r} f_B(v) \cdot [F_B(\lambda_{i+1}) - F_B(\lambda_i)]^{r-1}, & \text{for } \lambda_i \leq v < \lambda_{i+1}, i = 1, \dots, L-1, \end{cases} \quad (33)$$

where $F_B(v) \approx [F_Z(v)]^{MN}$ and $f_B(v) \approx MN f_Z(v) F_Z(v)^{MN-1}$. The PDF $f_Z(v)$ and the CDF $F_Z(v)$ are, respectively, given by (8) and (9). After substituting (33) into (32), we can obtain the approximation of the achievable throughput by the quantized feedback scheduling.

C. 1-bit Feedback

In the case of a quantization with 1-bit, each user feeds 1 or 0 back to the BS according to the threshold λ_1 , thereby, obtaining the minimal feedback load. In this case, (33) can be rewritten as

$$f_V(v) = \begin{cases} [F_B(\lambda_1)]^{K'-1} f_B(v), & 0 \leq v < \lambda_1 \\ \frac{1 - [F_B(\lambda_1)]^{K'}}{1 - F_B(\lambda_1)} f_B(v), & \lambda_1 \leq v. \end{cases} \quad (34)$$

Therefore, the average total throughput (32) is given by

$$E(R) = M[F_B(\lambda_1)]^{K'-1} \int_0^{\lambda_1} \log_2(1+v) f_B(v) dv + M \frac{1 - [F_B(\lambda_1)]^{K'}}{1 - F_B(\lambda_1)} \int_{\lambda_1}^{\infty} \log_2(1+v) f_B(v) dv, \quad (35)$$

where the approximation is obtained from $F_B(v) \approx [F_Z(v)]^{MN}$ and $f_B(v) \approx MN f_Z(v) F_Z(v)^{MN-1}$ with $K' =$

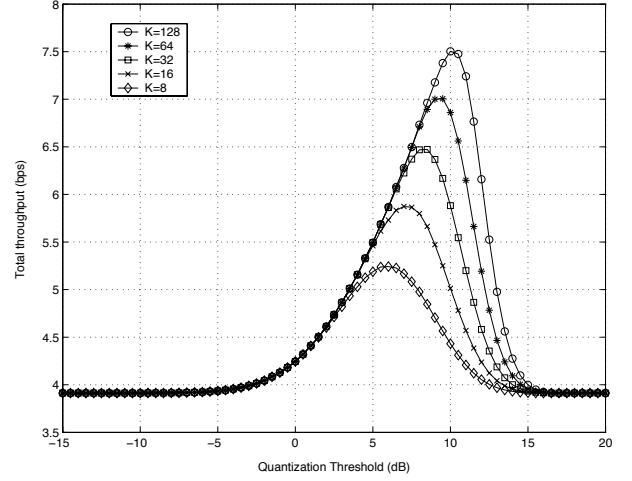


Fig. 10. Total throughput of the quantized feedback multiuser scheduling vs. threshold (1-bit quantization) for various number of users $K = 8, 16, 32, 64$ and 128. $M = 2$ and $N = 2$. SNR is 10 dB.

K/M . The PDF $f_Z(v)$ and the CDF $F_Z(v)$ are, respectively, given by (8) and (9).

When K goes to infinity (while $K' \rightarrow \infty$), it can be observed that $[F_Z(\lambda_1)]^{K'MN} \rightarrow 0$ for a fixed value $\lambda_1 < \infty$. Then, the throughput of (35) is derived as

$$\lim_{K \rightarrow \infty} E(R) \approx \frac{M^2 N}{1 - [F_Z(\lambda_1)]^{MN}} \times \int_{\lambda_1}^{\infty} \log_2(1+v) f_Z(v) [F_Z(v)]^{MN-1} dv \quad (36)$$

If the quantization threshold λ_1 is fixed, it can be seen from (36) that the total rate will be a constant. This implies that the multiuser diversity is lost in this case.

To validate the accuracy of (35), the quantized feedback scheduling is simulated for $M = 2, N = 2$ at SNR $\bar{\gamma} = 0, 5$ and 10 dB, respectively. The 1-bit quantization is used and the threshold λ_1 is always fixed as the value equal to $\bar{\gamma}$. Fig. 9 shows the simulation results and the numerical results of the average total throughput achieved by the quantized feedback scheduling. The performance comparison between the numerical and simulation results shows a good approximation of (35). Moreover, when K goes to infinity the throughput approaches a fixed value. It demonstrates that the multiuser diversity is lost, thereby, corroborating the analysis of (36).

For any given M, N and some $\bar{\gamma}$ in (35), it can be seen that the throughput is a function of λ_1 and K , simply denoted by $E(R) = f(K, \lambda_1)$. Hence, it is not optimal to fix the value of the threshold λ_1 for various K to enhance the throughput. Intuitively, an optimal λ_1 which maximizes throughput should depend on the value of K . To search for the optimal quantization threshold, we need to solve $\frac{\partial f(K, \lambda_1)}{\partial \lambda_1} = 0$. As seen from (35), a closed form solution to $\frac{\partial f(K, \lambda_1)}{\partial \lambda_1} = 0$ is not tractable. However, a numerical solution can be easily given. From (35), the numerical results of the achievable throughput with respect to the quantization threshold λ_1 are shown in Fig. 10 for various number of users, $K = 8, 16, 32, 64$ and 128 where $M = N = 2$ and $\bar{\gamma} = 10$ dB. It can be observed that the optimal values of the thresholds are 6, 7, 8.5, 9.5 and

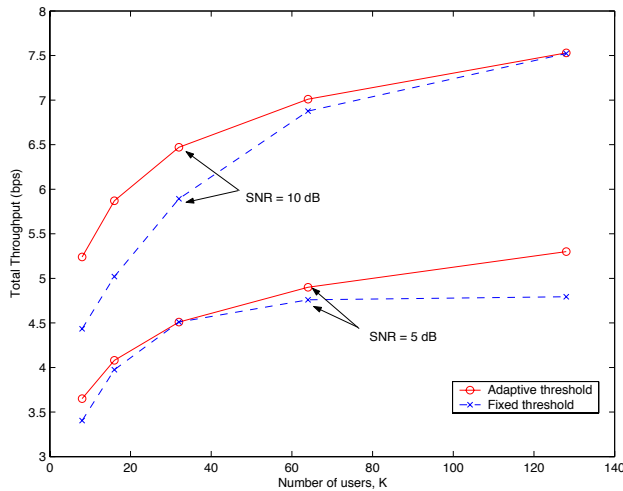


Fig. 11. Throughput comparison of the quantized feedback multiuser scheduling between using fixed quantization and adaptive (optimal) quantization thresholds. For the fixed threshold, λ is chosen the same as the SNR and irrespective of the number of users K . For the adaptive threshold, λ is optimized by maximizing the throughput for different K . $M = 2$ and $N = 2$.

10 dB, respectively. If the threshold is fixed to be 10 dB for all K , then only the case of $K = 128$ can achieve the maximum throughput. As we decrease K at $\lambda_1 = 10$ dB, the gap between the achievable throughput and the maximum throughput is increasing. Hence, we conclude that a fixed threshold is not optimal whereas the optimal threshold should be dependent on K for given M, N and SNR $\bar{\gamma}$.

In Fig. 11, the throughput of 1-bit quantized feedback scheduling with the optimal quantization threshold is compared with the one using the fixed threshold for $M = N = 2$ under 5 and 10 dB SNR. The optimal threshold is obtained by maximizing the throughput whereas the fixed threshold is chosen as the value equal to the SNR. This figure shows that the fixed threshold values achieve the maximum throughput only when $K = 32$ for 5 dB SNR ($\lambda_1 = 5$ dB) and $K = 128$ for 10 dB SNR ($\lambda_1 = 10$ dB), respectively.

VI. FURTHER RESULTS

Fig. 12 shows the performance comparison among the three scheduling algorithms proposed in this paper for $M = N = 2$ at 10 dB SNR. The threshold for selected feedback scheduling is chosen as 5 dB which can get an FLR 0.4 irrespective of K as shown in Section IV-D. For the 1-bit quantized feedback scheduling, the optimal quantization threshold is chosen from (35) adaptively for various K so as to achieve the maximal throughput. Fig. 12 shows that the selected feedback scheduling achieves the almost identical throughput as the full feedback scheduling for more than 32 users while obtaining a reduction of 60% in feedback load. The 1-bit quantized feedback scheduling has a lower throughput, but the multiuser diversity is obtained and suffers a loss around 0.6 bps compared with the full feedback scheduling.

VII. CONCLUSION

We have investigated several multiuser scheduling algorithms with limited feedback in MIMO broadcast channels.

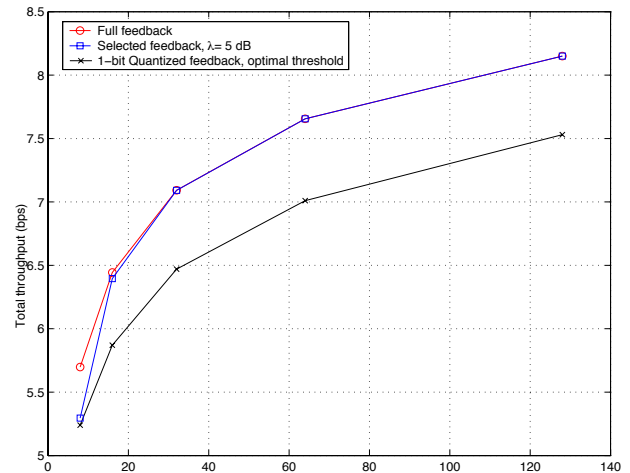


Fig. 12. Throughput comparison among the three scheduling algorithms: full feedback, selected feedback and quantized feedback. $M = N = 2$ and SNR is 10 dB.

Combined with spatial multiplexing and receive antenna selection, the proposed scheduling algorithms can achieve high multiuser diversity. It was shown that with user selection based on SINR, the feedback load can be greatly reduced with a negligible throughput loss. In particular, the scheduling with quantized feedback requires a smaller feedback load by transmitting a few bits of quantized SINR. As a special case, 1-bit quantization was shown to be very attractive as it can achieve multiuser diversity with minimal feedback load.

The aim of this paper was to consider broadcast scheduling with limited feedback of partial side information in MIMO channels. Other interesting topics remain open such as the investigation of broadcast scheduling algorithms for MIMO-OFDM systems as well as extensions to packet-based transmission where cross-layer design can be used to achieve the optimal throughput with quality-of-service (QoS) provisioning.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
- [2] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 684–702, June 2003.
- [3] S. M. Alamouti, "A simple transmit diversity technique for wireless communication," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [4] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, Nov. 1996.
- [5] S. Sfar, L. Dai, and K. B. Letaief, "Optimal diversity-multiplexing tradeoff with group detection for MIMO systems," *IEEE Trans. Commun.*, vol. 53, pp. 1178–1190, July 2005.
- [6] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE Int. Conf. Commun.*, June 1995, pp. 331–335.
- [7] K. B. Letaief and Y. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Commun. Mag.*, vol. 13, pp. 38–47, Aug. 2006.
- [8] W. Ajib and D. Haccoun, "An overview of scheduling algorithms in MIMO-based fourth-generation wireless systems," *IEEE Network*, pp. 43–48, Sep/Oct. 2005.

- [9] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.
- [10] R. Gozali, R. M. Buehrer, and B. D. Woerner, "The impact of multiuser diversity on space-time block coding," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 213–215, May 2003.
- [11] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.
- [12] E. G. Larsson, "On the combination of spatial diversity and multiuser diversity," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 517–519, Aug. 2004.
- [13] J. Chung, C.-S. Hwang, K. Kim and Y. K. Kim, "A random beamforming techniques in MIMO systems exploiting multiuser diversity," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 848–855, June 2003.
- [14] N. Sharma and L. H. Ozarow, "A study of opportunism for multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1804–1814, May 2005.
- [15] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1783–1794, May 2005.
- [16] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
- [17] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, USA, June 27–July 2, 2004, p. 174.
- [18] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1570–1580, Apr. 2005.
- [19] W. Yu, "Sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," *IEEE Trans. Inform. Theory*, vol. 52, pp. 754–759, Feb. 2006.
- [20] Z. Tu and R. S. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Commun. Lett.*, vol. 7, pp. 370–372, Aug. 2003.
- [21] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. Signal Processing*, vol. 53, pp. 3857–3868, Oct. 2005.
- [22] K. K. Wong, R. D. Murch, and K. B. Letaief, "A joint channel diagonalization for multi-user MIMO antenna systems," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 773–786, July 2003.
- [23] L.-U. Choi and R. D. Murch, "A transmit pre-processing technique for multi-user MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 20–24, Jan. 2004.
- [24] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication – Part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, pp. 195–202, Jan. 2005.
- [25] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "MIMO broadcast scheduling with limited channel state information," in *Proc. Annual Allerton Conf. on Commun., Control and Computing*, Sep. 2005.
- [26] A. Bayesteh and A. Khandani, "On the user selection for MIMO broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sept. 4–9, 2005, pp. 2325–2329.
- [27] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [28] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels," *IEEE Commun. Mag.*, pp. 54–59, Oct. 2004.
- [29] N. Jindal, "MIMO broadcast channels with finite rate feedback," in *Proc. IEEE Global Commun. Conf.*, St. Louis, MO, USA, Nov. 28–Dec. 2, 2005, vol. 3, pp. 1520–1524.
- [30] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.
- [31] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth," in *Proc. IEEE Int. Conf. Commun.*, June 20–24, 2004, vol. 1, pp. 234–238.
- [32] F. Floren, O. Edfors, and B.-A. Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme," in *Proc. IEEE Global Commun. Conf.*, 2003, pp. 497–501.
- [33] M. Johansson, "Benefits of multiuser diversity with limited feedback," in

Proc. IEEE Workshop on Signal Processing Adv. for Wireless Commun., June 2003, pp. 155–159.

- [34] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with only 1-bit feedback," in *Proc. IEEE Wireless Commun. and Networking Conf.*, Apr. 2005, pp. 978–983.

- [35] R. Zhang, J. M. Cioffi, and Y.-C. Liang, "Throughput comparison of wireless downlink transmission schemes with multiple antennas," in *Proc. IEEE Int. Conf. Commun.*, Seoul, Korea, May 16–20, 2005, vol. 4, pp. 2700–2704.

- [36] H. A. David, *Order Statistics*, 2nd ed. New York: Wiley, 1981.



Wei Zhang (M'06) received his B. Eng. degree in mechanical engineering from Jilin University of Technology, Changchun, China, his M. Eng. degree in communications engineering from Jilin University, Changchun, China, and his Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, in 1998, 2002, and 2005, respectively.

From 1998 to 1999, he was Research Assistant with the Institute of Oceanographic Instrumentation, Shandong Academy of Sciences, Tsingdao, China.

He was Visiting Student at the Department of Electrical and Computer Engineering, University of Delaware, USA, in 2004. He is presently Postdoctoral Fellow at the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. His current research interests include multiuser MIMO, cooperative diversity, and cognitive radio.



Khaled Ben Letaief (S'85-M'86-SM'97-F'03) received the BS degree with distinction in Electrical Engineering from Purdue University at West Lafayette, Indiana, USA, in December 1984. He received the MS and Ph.D. Degrees in Electrical Engineering from Purdue University, in August 1986, and May 1990, respectively. From January 1985 and as a Graduate Instructor in the School of Electrical Engineering at Purdue University, he has taught courses in communications and electronics.

From 1990 to 1993, he was a faculty member at the University of Melbourne, Australia. Since 1993, he has been with the Hong Kong University of Science & Technology where he is currently Chair Professor and Head of the Electronic and Computer Engineering Department. He is also the Director of the Hong Kong Telecom Institute of Information Technology as well as the Director of the Center for Wireless Information Technology. His current research interests include wireless and mobile networks, Broadband wireless access, OFDM, CDMA, Cooperative networks, Cognitive radio, MIMO, and Beyond 3G systems. In these areas, he has published over 300 journal and conference papers and given invited keynote talks as well as courses all over the world.

Dr. Letaief served as consultants for different organizations and is the founding Editor-in-Chief of the IEEE Transactions on Wireless Communications. He has served on the editorial board of other prestigious journals including the IEEE Journal on Selected Areas in Communications - Wireless Series (as Editor-in-Chief). He has been involved in organizing a number of major international conferences and events. These include serving as the Co-Technical Program Chair of the 2004 IEEE International Conference on Communications, Circuits and Systems, ICCCS'04; General Chair of the 2007 IEEE Wireless Communications and Networking Conference, WCNC'07; as well as the Technical Program Co-Chair of the 2008 IEEE International Conference on Communication, ICC'08.

In addition to his active research and professional activities, Professor Letaief has been a dedicated teacher committed to excellence in teaching and scholarship. He received the Mangoon Teaching Award from Purdue University in 1990; the Teaching Excellence Appreciation Award by the School of Engineering at HKUST (4 times); and the Michael G. Gale Medal for Distinguished Teaching (Highest university-wide teaching award and only one recipient/year is honored for his/her contributions).

He is a Fellow of IEEE, an elected member of the IEEE Communications Society Board of Governors, and an IEEE Distinguished lecturer of the IEEE Communications Society. He also served as the Chair of the IEEE Communications Society Technical Committee on Personal Communications.