

# QoS-driven asynchronous uplink subchannel allocation algorithms for space-time OFDM-CDMA systems in wireless networks

Xi Zhang · Jia Tang

Published online: 8 May 2006  
© Springer Science + Business Media, LLC 2006

**Abstract** In order to support the diverse Quality of Service (QoS) requirements for differentiated data applications in broadband wireless networks, advanced techniques such as space-time coding (STC) and orthogonal frequency division multiplexing (OFDM) are implemented at the physical layer. However, the employment of such techniques evidently affects the subchannel-allocation algorithms at the medium access control (MAC) layer. In this paper, we propose the QoS-driven cross-layer subchannel-allocation algorithms for data transmissions over asynchronous uplink space-time OFDM-CDMA wireless networks. We mainly focus on QoS requirements of maximizing the best-effort throughput and proportional bandwidth fairness, while minimizing the upper-bound of scheduling delay. Our extensive simulations show that the proposed infrastructure and algorithms can achieve high bandwidth fairness and system throughput while reducing scheduling delay over wireless networks.

**Keywords** Quality of Service (QoS) · Cross-layer design and optimization · Wireless networks · Fairness · Throughput

## 1. Introduction

The increasing demand for wireless network services such as the wireless Internet access, mobile computing, and wireless

communications motivates an unprecedented revolution in the wireless broadband access [1, 2]. This presents great challenges in designing the wireless networks since the wireless channel has the significant impact on supporting the various quality of service (QoS) requirements for different users.

A number of promising schemes are developed at the physical layer to overcome the impact of wireless channels. Among them, space-time (ST) processing using multiple-input-multiple-output (MIMO) architecture emerges as one of the important technical breakthroughs in wireless communications [3–7]. Besides, the combination of the widely employed code division multiple access (CDMA) with orthogonal frequency division multiplexing (OFDM), called OFDM-CDMA, takes the advantages of these two techniques and also receives a great deal of research efforts [7–11]. Clearly, employment of the integrated design combining space-time processing and OFDM-CDMA can achieve the integrated diversities from spatial, temporal, frequency, and code domains, which will result in significant improvements in supporting QoS for differentiated data users with diverse QoS requirements over wireless networks.

Space-Time OFDM-CDMA provides us with not only the diversities, but also the multiple access control, such that multiple users can be assigned into a single subchannel, where they distinct themselves from each other by different signature sequences. As a result, how to allocate the resources efficiently in satisfying the various QoS requirements becomes increasingly critical. While there have already been a large body of literature on both space-time processing and OFDM-CDMA, the impact of such architectures on resource allocations at MAC layer, its bandwidth fairness, throughput, delay analysis, and cross-layer optimizations, have received relatively much less attention. Therefore, it is important to develop a cross-layer scheme to integrate the resource

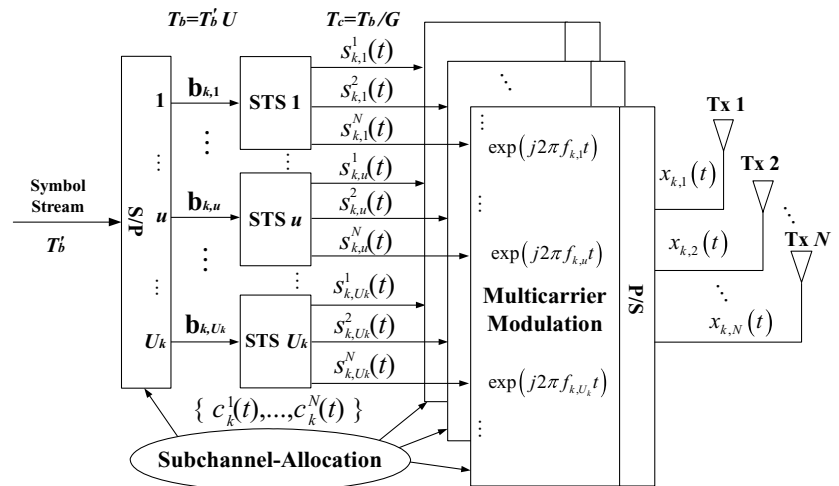
---

The research reported in this paper was supported in part by the National Science Foundation CAREER Award under Grant ECS-0348694.

---

X. Zhang (✉) · J. Tang  
Networking and Information Systems Laboratory, Department of  
Electrical Engineering, Texas A&M University, College Station,  
TX 77843 USA  
e-mail: {xizhang, jtang}@ece.tamu.edu

**Fig. 1** Block diagram of the  $k$ th mobile user's uplink transmitter



allocation at the MAC layer and the multi-antenna infrastructure implemented at the physical layer.

The subcarrier allocation for OFDM system has been studied in, e.g., [12–14]. However, the previous work mainly focuses on TDMA-, FDMA-, or SDMA-based systems. In this paper, we propose the QoS-driven cross-layer subchannel-allocation algorithms for asynchronous uplink space-time OFDM-CDMA systems targeting at differentiated data transmission applications. The QoS requirements that we focus on include transmission reliability, bandwidth fairness, system throughput, and scheduling delay. Specifically, the proposed algorithms are based on the delay-fairness-driven scheduling, proportional bandwidth fairness, and best-effort throughput. Note that our proposed proportional fairness differs slightly from the definition used in [15], where the assigned bandwidth is proportional to the long-term average rate considering the soft-delay constraint. In our scheme, the bandwidth is proportional to the service-priorities of the differentiated services. We also conduct extensive simulations to evaluate the performance of the proposed algorithms. Our simulation results show that the proposed infrastructures and algorithms can significantly improve the bandwidth fairness and system throughput, while achieving more efficient resource allocations over wireless networks.

The paper is organized as follows. Section 2 describes the space-time OFDM-CDMA system model. Section 3 proposes the subchannel-allocation algorithms. Section 4 evaluates and compares the various performance metrics through simulations and numerical solutions. The paper concludes with Section 5.

## 2. System model

We consider the BPSK modulation-based uplink from a mobile user to the basestation in a packet-cellular network

with  $M$  antennas at the basestation (BS) and  $N$  antennas at each mobile user. Let  $K$  denote the total number of mobile users and  $U$  the total number of subcarriers or subchannels,<sup>1</sup> which are to be assigned to the  $K$  mobile users. The index set of all  $K$  users is defined as  $\Omega \triangleq \{1, 2, \dots, K\}$  and the  $U$  subcarrier frequencies are denoted by the set of  $\{f_u\}_{u=1}^U$ . In addition, at any instantaneous time point, only the users which have been assigned with bandwidth resources can transmit data packets. These users assigned with bandwidth resources are defined as *active* users. Denote the index set of active users by  $\Phi$ . Clearly, the set of the active users is a subset of the all-user set  $\Omega$ , i.e.,  $\Phi \subseteq \Omega$ .

### 2.1. Mobile uplink transmitter model

The  $k$ th mobile user's transmitter structure of our proposed space-time OFDM-CDMA system is shown in Fig. 1. In this paper, we consider the asynchronous uplink, where different users transmit data asynchronously to the basestation, and thus we cannot use the synchronous model as used for downlink from the basestation. As shown in Fig. 1, using the Serial-to-Parallel (S/P) converter, a block of  $U_k \times N$  BPSK symbols each with bit duration of  $T_b'$  is converted to  $U_k$  parallel sub-streams. The value  $U_k (U_k \leq U)$  is determined by our proposed subchannel-allocation algorithms, which will be described in Section III with more details. Each of  $U_k$  sub-streams consists of  $N$  bits, which are denoted by  $\mathbf{b}_{k,u} = (b_{k,u}^1 \ b_{k,u}^2 \ \dots \ b_{k,u}^N)^T$ , where  $(\cdot)^T$  represents the transpose of  $(\cdot)$  and  $u \in \{1, 2, \dots, U_k\}$ . The bit duration  $T_b$  after S/P conversion becomes  $T_b = T_b' U$ . Then, each  $\mathbf{b}_{k,u}$  is space-time spread (STS) [4] using the spreading code given by  $\mathbf{c}_k = (c_k^0 \ c_k^1 \ \dots \ c_k^{G-1})^T$ , where  $c_k^g \in \{\pm 1\}$ ,

<sup>1</sup> We use the terms “subchannel” and “subcarrier” interchangeably in the following discussions.

$g \in \{0, 1, \dots, G - 1\}$  and  $G$  denotes the spreading gain of the code. The chip duration  $T_c$  of the spreading code satisfies  $T_c = T_b/G = T'_b U/G$ . The waveform expression  $c_k(t)$  of the spreading code corresponding to  $\mathbf{c}_k$  is determined by

$$c_k(t) = \frac{1}{\sqrt{G}} \sum_{g=0}^{G-1} c_k^g p(t - gT_c), \quad 0 \leq t < T_b \tag{1}$$

where  $p(t)$  is a normalized rectangular chip waveform which has the finite duration  $[0, T_c]$ . In this paper, we focus on a specific subset of the general STS schemes investigated in [4], by which the  $N$ -bit data is coded, spread, and allocated to  $N$  transmit antennas, and then transmitted by  $N$  time intervals. This kind of STS schemes can provide the maximal transmit diversity without demanding extra spreading codes and thus be considered as *attractive* schemes [7]. Denote the corresponding space-time block coding square matrix of the  $u$ th sub-stream by

$$\mathbf{B}_{k,u} = \begin{pmatrix} b_{k,u}^{1,1} & b_{k,u}^{1,2} & \dots & b_{k,u}^{1,N} \\ b_{k,u}^{2,1} & b_{k,u}^{2,2} & \dots & b_{k,u}^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,u}^{N,1} & b_{k,u}^{N,2} & \dots & b_{k,u}^{N,N} \end{pmatrix} \begin{matrix} \rightarrow \text{space} \\ \downarrow \text{time} \end{matrix} \tag{2}$$

where the rows and columns of matrix  $\mathbf{B}_{k,u}$  consist of  $\mathbf{b}_{k,u}$  with different signs and sequences according to the orthogonal-design principles [5]. The spreading code vector  $\mathbf{c}_k(t)$  used for STS can be expressed as

$$\mathbf{c}_k(t) = (c_k^1(t) \ c_k^2(t) \ \dots \ c_k^N(t)) \tag{3}$$

where  $c_k^i(t)$ ,  $i \in \{1, 2, \dots, N\}$  has the finite duration  $[0, NT_b)$ , which is given by

$$c_k^i(t) = \begin{cases} c_k(t - (i - 1)T_b), & (i - 1)T_b \leq t < iT_b \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Using Eqs. (2) and (3), STS can be expressed as

$$\mathbf{s}_{k,u}(t) \triangleq (s_{k,u}^1(t) \ s_{k,u}^2(t) \ \dots \ s_{k,u}^N(t)) = \mathbf{c}_k(t)\mathbf{B}_{k,u}. \tag{5}$$

Following STS, the  $U_k$  data streams of each antenna are transmitted simultaneously by modulating  $U_k$  different subcarriers, which can be implemented by the operation of inverse fast Fourier transform (IFFT). The frequency spacing  $\Delta$  between any of the adjacent subcarriers  $\{f_u\}_{u=1}^U$  satisfies  $\Delta = 1/T_c$ , guaranteeing the orthogonal subcarrier condition. The selection of which  $U_k$  out of  $U$  subcarriers are used to

transmit data is also determined by our proposed subchannel-allocation algorithms. Denote the  $U_k$  subcarrier central frequencies assigned to the  $k$ th user by  $\{f_{k,u}\}_{u=1}^{U_k}$ , the transmitted signal  $x_{k,n}(t)$  from the  $n$ th transmit antenna of the  $k$ th user to the basestation within a block-interval  $[\tau_k, \tau_k + NT_b)$  can be expressed by

$$x_{k,n}(t) = \sqrt{\frac{P}{NU}} \sum_{u=1}^{U_k} \sum_{i=1}^N b_{k,u}^{i,n} c_k^i(t - \tau_k) e^{j2\pi f_{k,u}(t - \tau_k)} \tag{6}$$

where  $\tau_k$  represents the transmission delay of the  $k$ th user which satisfies  $0 \leq \tau_k < T_b$ ;  $P$  denotes the maximum transmission power, which can be achieved when  $U_k = U$ ; the coefficient  $\sqrt{P/(NU)}$  indicates that the maximum transmission power is independent of the total numbers of transmit antennas and subcarriers;  $b_{k,u}^{i,n}$  is given by Eq. (2) and  $c_k^i(t)$  is given by Eq. (4), respectively. Clearly, the transmission throughput is proportional to the assigned number of subcarriers. The larger the number  $U_k$  of subcarriers assigned to the  $k$ th user, the higher the throughput can be achieved.

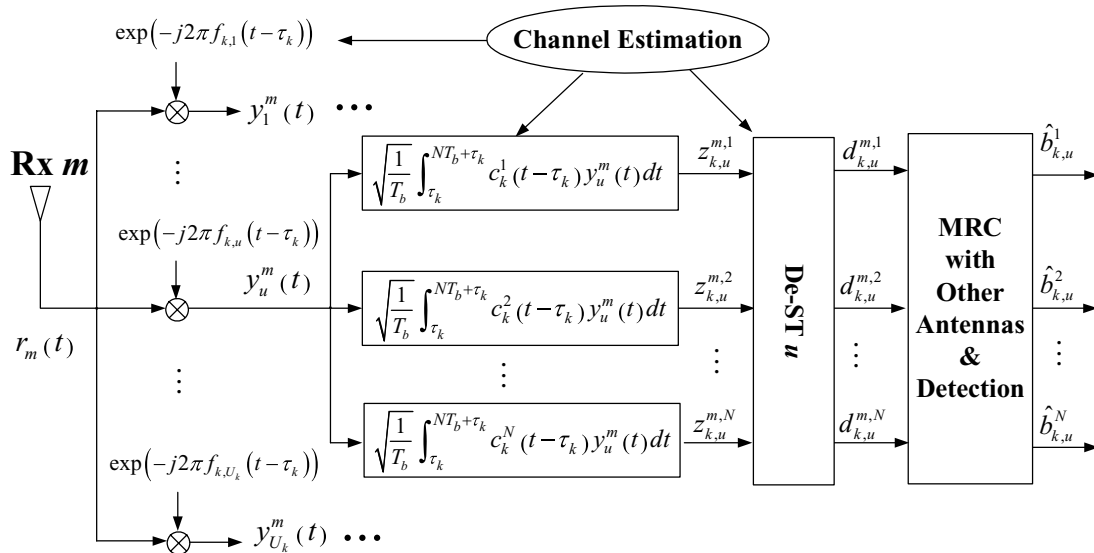
### 2.2. Uplink channel model

We assume that the Rayleigh fading channel is frequency-selective, but the delay-spreads  $T_m$  of the channel satisfy  $T_m \ll T_c$  such that each subchannel conforms to the flat fading. In addition, the channel is assumed to be *quasi-static*, i.e., the fading coefficients are invariant over a block-interval and vary from one block to another. Thus, during each block-interval, the fading coefficient of the  $u$ th subcarrier  $h_{k,u}^{n,m}(t)$ , between  $n$ th transmit antenna of the  $k$ th user and the  $m$ th receive antenna of the basestation, can be denoted by  $h_{k,u}^{n,m}[i]$ , where  $i \in \{1, 2, \dots\}$  is the discrete time index for the  $i$ th block interval. The time-varying channel can be modeled by an auto-regressive (AR) process [11] as follows

$$h_{k,u}^{n,m}[i] = \xi_k h_{k,u}^{n,m}[i - 1] + v_{k,u}^{n,m}[i] \tag{7}$$

where  $\xi_k$  is determined by the  $k$ th user's Doppler velocity and  $v_{k,u}^{n,m}[i]$  is a zero-mean *independent identically distributed* (i.i.d.) complex-Gaussian variable. In Section 2 and Section 3 we will focus on the discussion within a block interval. Therefore, we drop the time index  $i$  for convenience. We assume that  $\{h_{k,u}^{n,m} \mid \forall k, u, m, n\}$  are independent<sup>2</sup> identically distributed (i.i.d.) complex-Gaussian variables with

<sup>2</sup> Theoretically, the coefficients between different subcarriers are not independent. However, this independence assumption is valid if we employ frequency-interleaving operations [7, 11]. In this paper, we omit the frequency-interleaving to simplify the presentation.



**Fig. 2** Uplink data-/signal-flow received by the  $m$ th receive antenna at the base station for the  $k$ th user

zero-mean and variance of  $\sigma^2$  per dimension, i.e.,

$$\sigma = E(|h_{k,u}^{n,m}|^2) \tag{8}$$

Assuming perfect power control, the received signal  $r_m(t)$  received at the  $m$ th receive antenna at the base station, as shown in Fig. 2, by

$$r_m(t) = \sqrt{\frac{P}{NU}} \sum_{k \in \Phi} \sum_{n=1}^N \sum_{u=1}^{U_k} \sum_{i=1}^N h_{k,u}^{n,m} b_{k,u}^{i,n} c_k^i(t - \tau_k) \cdot e^{j2\pi f_{k,u}(t-\tau_k)} + w_m(t) \tag{9}$$

where  $w_m(t)$  denotes the complex Additive White Gaussian Noise (AWGN) at the  $m$ th receive antenna with zero-mean and double-sided power-spectral density of  $N_0/2$ . Demodulating the received signal  $r_m(t)$  given by Eq. (9) and expressing the multiple demodulated signals in a matrix form, the obtained complex low-pass signal received at the  $m$ th receive antenna through the  $u$ th subchannel, denoted by  $y_u^m(t)$  for  $u \in \{1, 2, \dots, U_k\}$  (see Fig. 2), can be expressed as

$$y_u^m(t) = \sqrt{\frac{P}{NU}} \sum_{k \in \Phi} \mathbf{c}_k(t - \tau_k) \mathbf{B}_{k,u} \mathbf{h}_{k,u}^m + N_u^m(t) = \sqrt{\frac{P}{NU}} \sum_{k \in \Phi} \mathbf{c}_k(t - \tau_k) \mathbf{H}_{k,u}^m \mathbf{b}_{k,u} + N_u^m(t) \tag{10}$$

where  $\mathbf{h}_{k,u}^m = (h_{k,u}^{1,m} h_{k,u}^{2,m} \dots h_{k,u}^{N,m})^T$  denotes the channel vector from the mobile user to the base station and the matrix  $\mathbf{H}_{k,u}^m$

has the form as follows:

$$\mathbf{H}_{k,u}^m = \begin{pmatrix} h_{k,u,m}^{1,1} & h_{k,u,m}^{1,2} & \dots & h_{k,u,m}^{1,N} \\ h_{k,u,m}^{2,1} & h_{k,u,m}^{2,2} & \dots & h_{k,u,m}^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k,u,m}^{N,1} & h_{k,u,m}^{N,2} & \dots & h_{k,u,m}^{N,N} \end{pmatrix} \tag{11}$$

where the rows and the columns of  $\mathbf{H}_{k,u}^m$  consist of  $\mathbf{h}_{k,u}^m$  with different signs and sequences according to the orthogonal-design principles [5], i.e.,

$$\mathbf{B}_{k,u} \mathbf{h}_{k,u}^m = \mathbf{H}_{k,u}^m \mathbf{b}_{k,u}. \tag{12}$$

### 2.3. Uplink receiver model at the base station

The schematic for the  $m$ th receive antenna at the base station of our proposed space-time OFDM-CDMA system is shown in Fig. 2, where we focus on the decoding scheme for the  $k$ th user within a block interval  $[\tau_k, \tau_k + NT_b)$ . In addition, we assume that the channel state information (CSI) and the transmission delay of each user can be perfectly estimated by the base station.

Performing the inverse operation of the transmitter, the received signal  $r_m(t)$  at the  $m$ th antenna is split into  $U_k$  sub-streams by demodulating  $U_k$  different sub-carriers  $\{e^{-j2\pi f_{k,u}(t-\tau_k)}\}_{u=1}^{U_k}$ . Then, each sub-stream correlates with the  $k$ th user’s referenced waveforms  $\{c_k^i(t - \tau_k)\}_{i=1}^N$  during  $[\tau_k, \tau_k + NT_b)$  to obtain correlation outputs  $\mathbf{z}_{k,u}^m = (z_{k,u}^{m,1} z_{k,u}^{m,2} \dots z_{k,u}^{m,N})^T$ . Then, the space-time decoding (De-ST) is employed to obtain  $N$  decision variables

$\mathbf{d}_{k,u}^m = (d_{k,u}^{m,1} d_{k,u}^{m,2} \dots d_{k,u}^{m,N})^T$ , which can be expressed as

$$\mathbf{d}_{k,u}^m = (\mathbf{H}_{k,u}^m)^* \mathbf{z}_{k,u}^m \tag{13}$$

where  $(\cdot)^*$  denotes the conjugate transpose of  $(\cdot)$ ,  $\mathbf{H}_{k,u}^m$  is given by Eq. (11), and  $\mathbf{d}_{k,u}^m$  corresponds to the original transmitted  $N$  bits expressed by  $\mathbf{b}_{k,u} = (b_{k,u}^1 b_{k,u}^2 \dots b_{k,u}^N)^T$ . Following space-time decoding, all decision variables  $\{\mathbf{d}_{k,u}^m\}_{m=1}^M$  obtained from  $M$  receive antennas are combined together as follows:

$$\hat{\mathbf{d}}_{k,u} = \sum_{m=1}^M \mathbf{d}_{k,u}^m, \quad \forall k \in \Phi, \forall u = 1, 2, \dots, U_k \tag{14}$$

which represents the procedure of Maximum Ratio Combining (MRC). Based upon decision variables given in Eq. (14), the receiver makes the decisions of the transmitted bits (see Fig. 2) by

$$\hat{\mathbf{b}}_{k,u} \triangleq (\hat{b}_{k,u}^1 \hat{b}_{k,u}^2 \dots \hat{b}_{k,u}^N)^T = \text{sgn}[\text{Re}(\hat{\mathbf{d}}_{k,u})] \tag{15}$$

where  $\text{sgn}(\cdot)$  is the signum function and  $\text{Re}(\cdot)$  denotes the real part of  $(\cdot)$ . Denote the index-set of active users allocated in the  $u$ th subchannel by  $\Phi_u$ . It is easy to see that

$$\Phi = \bigcup_{u=1}^U \Phi_u. \tag{16}$$

To derive the SINR of the decoded signals received from the  $k$ th user ( $k \in \Phi_u$ ) through the  $u$ th subchannel, we apply the property of the orthogonal-design principles [5] such that

$$(\mathbf{H}_{k,u}^m)^* \mathbf{H}_{k,u}^m = (\mathbf{h}_{k,u}^m)^* \mathbf{h}_{k,u}^m \mathbf{I}_N = \left( \sum_{n=1}^N |h_{k,u}^{m,n}|^2 \right) \mathbf{I}_N \tag{17}$$

where  $\mathbf{I}_N$  represents a  $N \times N$  unity matrix. Then, from  $y_u^m(t)$  given by Eq. (10), the signal energy associated with the decoded signal received from the  $k$ th user through the  $u$ th subchannel, which is the summation of the signal gains over all  $(M \times N)$  transmit-to-receive antenna channels, is determined by

$$\left| \sqrt{\frac{PT_b}{NU}} \sum_{m=1}^M (\mathbf{h}_{k,u}^m)^* \mathbf{h}_{k,u}^m \right|^2 = \frac{PT_b}{NU} \left( \sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{m,n}|^2 \right)^2 \tag{18}$$

The variance of the interference associated with the decoded signal received from the  $k$ th user through the  $u$ th subchannel

can be expressed as:

$$\left( \frac{PT_b}{NU} \right) \sum_{m=1}^M \left( \frac{\rho_{k,l}^2(\tau_{k,l})}{N} \right) \sum_{l \in \Phi_u, l \neq k} \|(\mathbf{H}_{k,u}^m)^* \mathbf{H}_{l,u}^m\|_F^2 \tag{19}$$

where  $\|\cdot\|_F$  represents the Frobenius matrix norm [16],  $\tau_{k,l} = (\tau_k - \tau_l)$ , and  $\rho_{k,l}(\tau_{k,l})$  denotes the correlation factor between spreading codes  $c_k(t - \tau_k)$  and  $c_l(t - \tau_l)$ . The variance of the noise associated with the decoded signal received from the  $k$ th user through the  $u$ th subchannel is given by

$$\sum_{m=1}^M (\mathbf{h}_{k,u}^m)^* \mathbf{h}_{k,u}^m N_0 = \left( \sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{m,n}|^2 \right) N_0. \tag{20}$$

Using Eqs. (18) through (20), we obtain the SINR of decoded signals for the  $k$ th user ( $k \in \Phi_u$ ) at the  $u$ th subchannel as expressed by Eq. (21),

$$\text{SINR}_{k,u} = \frac{\left( \frac{PT_b}{NU} \right) \left( \sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{m,n}|^2 \right)^2}{\left( \frac{PT_b}{NU} \right) \sum_{m=1}^M \sum_{l \in \Phi_u, l \neq k} \rho_{k,l}^2(\tau_{k,l}) \|(\mathbf{H}_{k,u}^m)^* \mathbf{H}_{l,u}^m\|_F^2 + \left( \sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{m,n}|^2 \right) N_0} \tag{21}$$

### 3. Subchannel-allocation algorithms

In order to design the subchannel-allocation algorithms for data transmissions, we mainly focus on QoS requirements for guaranteeing *transmission reliability*, maximizing *system throughput*, optimizing *proportional bandwidth fairness*, and minimizing *maximum scheduling delay*. In the rest of this section, we will discuss each of these issues in detail.

#### 3.1. Optimizing the SINR-threshold

Due to the nature of CDMA technique employed in our system, subchannel can be reused by multiple users. This introduces the following tradeoff. On one hand, in order to increase the system throughput, we need to assign as many users as possible into a single subchannel. The larger the number of users assigned into a subchannel, the higher the system throughput can be achieved. On the other hand, each user experiences co-subchannel interferences from other users within the same subchannel. When the number of users within the same subchannel becomes large, the interferences increase and users' SINRs decrease. As a result, the bit-error rate (BER) at the physical layer and the corresponding packet-loss/error rate at higher network-protocol layers increases. To characterize the tradeoff between error rate and throughput, we need to introduce a pre-determined

SINR threshold denoted by  $\gamma$ . To ensure the transmission reliability, each user that is assigned with the  $u$ th subcarrier,  $u \in \{1, 2, \dots, U\}$ , must satisfy  $\text{SINR} > \gamma$  requirement, which can be expressed as

$$\text{SINR}_{k,u} \geq \gamma, \quad \forall k \in \Phi_u. \quad (22)$$

In order to guarantee the transmission reliability QoS for data transmissions, the upper-layer protocol must employ some error-control methods, e.g., ARQ based protocol, to recover the data error/loss caused at the physical layer. When we choose a too small SINR threshold  $\gamma$ , the system throughput in terms of the number of users per subcarrier is high, but the number of retransmissions will be increased due to random loss; when we choose a too large SINR threshold  $\gamma$ , the number of retransmissions will be reduced, but the system throughput drops because more subchannel admission requests are rejected. This implies that there is the optimal SINR threshold  $\gamma$  that can maximize the system *goodput*, which is defined as the rate the packets are transmitted without retransmissions. When satisfying  $\text{SINR} \geq \gamma$  requirement, the BER  $P_b$  of physical layer using BPSK modulation is upper-bounded by

$$P_b \leq Q\left(\sqrt{2\gamma}\right) \quad (23)$$

where  $Q(\cdot)$  denotes the Q-function defined by

$$Q(x) \triangleq \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (24)$$

The corresponding packet error rate  $P_{pkt}$  is upper-bounded by

$$P_{pkt} \leq 1 - (1 - P_b)^L \quad (25)$$

where  $L$  denotes the packet size and we assume that the bit-losses/errors are independent with the probability equal to  $P_b$ . When applying Selective-Repeat ARQ, the system *goodput*  $\tilde{R}$  can be expressed as

$$\tilde{R} = \bar{R}(1 - P_{pkt}) \quad (26)$$

where  $\bar{R}$  denotes the system throughput. Finally, the optimal SINR threshold  $\gamma_{opt}$  can be derived by

$$\gamma_{opt} = \arg \max_{\gamma} \{\tilde{R}\}. \quad (27)$$

Note that we focus on uncoded system in this paper. For the coded system, the approach for finding the optimal threshold is similar but the threshold is smaller. Thus, more users can be assigned into the same subcarrier.

### 3.2. The proportional bandwidth fairness

For data transmission over wireless networks, different mobile users usually have different bandwidth capacities due to hardware or power facilities and service priorities due to the pricing schemes. This differentiated QoS in bandwidth can be achieved by employing the fairness-design criterion. The fairness problems have been widely studied in literatures [15, 17–19]. In this paper, we develop our algorithms based on the *proportional bandwidth fairness*, where the allocated bandwidth to different users is proportional to their different bandwidth capacities and service priorities. Thus, the users with the higher bandwidth capacities or priorities, which are characterized by their maximum bandwidth capacities  $U_k^{\max}$  (the maximum number of subcarriers assigned to the  $k$ th user), will receive the larger bandwidth-resource allocations. Note that our proposed proportional fairness is slightly different from the definition used in [15], where the assigned bandwidth is proportional to the long-term average rate considering the soft-delay constraint.

Let the  $k$ th user's maximum bandwidth requirement be  $U_k^{\max}$  ( $U_k^{\max} \leq U$ ) in terms of the number of subchannels assigned to the  $k$ th user. To formally define the fairness factor, we introduce the *aggressive factor*  $\alpha_k$  by:

$$\alpha_k \triangleq \frac{U_k}{U_k^{\max}}, \quad k \in \Omega \quad (28)$$

where  $0 \leq \alpha_k \leq 1$ . Then, we can define *fairness factor*  $\phi$  [17] by:

$$\phi \triangleq \frac{(\sum_{k \in \Omega} \alpha_k)^2}{K \sum_{k \in \Omega} \alpha_k^2} \quad (29)$$

where the cardinality of  $|\Omega| = K$  (the total number of users) and  $\alpha_k$  is given by Eq. (28). It is easy to see that  $0 \leq \phi \leq 1$ . The perfectly fair allocation has the maximum fairness factor  $\phi = 1$ , which is attained when  $\alpha_k = \alpha_j, \forall k \neq j$ , i.e., the bandwidth allocated to all users are equally proportional to their maximum bandwidth capacity  $U_k^{\max}$ . The worst bandwidth-allocation fairness has  $\phi = 1/K \rightarrow 0$  as  $K \rightarrow \infty$  when only one user occupies all the bandwidth resources. Obviously, the more the number of inactive users, the lower the bandwidth fairness attained. It is worth noting that when all  $\{U_k^{\max}\}_{k=1}^K$  are equal, the proposed scheme reduces to the conventional max-min fairness.

Our proposed subchannel-allocation algorithm can be divided into three steps, which are called *allocation-scheduling*, *initial allocation*, and *dynamic allocation*, respectively. In Section 3.3 and 3.4, we first describe our initial and dynamic allocation algorithms. Then, in Section 3.5 we

**Fig. 3** Pseudocode of initial subchannel-allocation algorithms

```

00. Procedure: Initial Allocation Algorithms;
01.  $\Phi := \emptyset;$  ! The initial value of the index-set of the active users
02.  $\{\Phi_u\}_{u=1}^U := \emptyset;$  ! The initial value of the index-set of the active users in each subcarrier
03. for  $j := 1$  to  $K$  do ! The sequence is based on deadline-fairness-driven scheduling
04.   for  $u := 1$  to  $U$  do ! Sequentially test all subcarriers
05.     begin Joining-Test for the  $j$ th user at the  $u$ th subcarrier; ! Using Eq. (30)
06.       if Eq. (30) is satisfied then
07.         begin  $\Phi := \Phi + \{j\};$  ! The  $j$ th user becomes an active user
08.            $\Phi_u := \Phi_u + \{j\};$  ! The  $j$ th user becomes an active user at the  $u$ th subcarrier
09.         break; ! Test the next user
10.       end;
11.   end;
    
```

present our allocation-scheduling scheme. All algorithms are exerted each time the system receives its CSI information.

### 3.3. Initial allocation algorithm

Let the  $j$ th user ( $j \in \Omega, j \notin \Phi_u$ ) be the candidate which attempts to transmit data using the  $u$ th subcarrier. Since the basestation knows the information and the statistical characteristics about the channel, we can pre-compute the SINR of decoding signals for each user at the  $u$ th subcarrier using Eq. (21). The  $j$ th user is admitted to use the  $u$ th subcarrier if and only if

$$\begin{cases} \text{SINR}_{j,u} \geq \gamma \\ \text{SINR}_{k,u} \geq \gamma, \forall k \in \Phi_u \end{cases} \quad (30)$$

where  $\text{SINR}_{j,u}$  and  $\text{SINR}_{k,u}$ , as the special cases of Eq. (21), are then derived and expressed by Eqs. (31) and (32), respectively,

$$\text{SINR}_{j,u} = \frac{\left(\frac{PT_b}{NU}\right) \left(\sum_{m=1}^M \sum_{n=1}^N |h_{j,u}^{n,m}|^2\right)}{\left(\frac{PT_b}{N^2U}\right) \sum_{m=1}^M \sum_{k \in \Phi_u} \rho_{k,j}^2(\tau_{k,j}) \left\| (\mathbf{H}_{k,u}^m)^* \mathbf{H}_{j,u}^m \right\|_F^2 + \left(\sum_{m=1}^M \sum_{n=1}^N |h_{j,u}^{n,m}|^2\right) N_0} \quad (31)$$

$$\text{SINR}_{k,u} = \frac{\left(\frac{PT_b}{NU}\right) \left(\sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{n,m}|^2\right)}{\left(\frac{PT_b}{N^2U}\right) \sum_{m=1}^M \sum_{l \in \Phi_u \cup \{j\}, l \neq k} \rho_{k,l}^2(\tau_{k,l}) \left\| (\mathbf{H}_{k,u}^m)^* \mathbf{H}_{l,u}^m \right\|_F^2 + \left(\sum_{m=1}^M \sum_{n=1}^N |h_{k,u}^{n,m}|^2\right) N_0} \quad (32)$$

If Eq. (30) is satisfied, the  $j$ th user is qualified to be assigned to the  $u$ th subcarrier. Then, it becomes a new active user in  $\Phi_u$ . Otherwise, if any of the SINR in Eq. (30) is lower than the threshold  $\gamma$ , the  $j$ th user is rejected to use the  $u$ th subcarrier.

Our initial allocation algorithm attempts to “activate” as many users as possible under the constraint that these users’

$\text{SINR} > \gamma$ . Figure 3 illustrates the pseudocode of the proposed initial allocation algorithm. As shown in Fig. 3, the  $U$  subcarriers are sequentially tested for each user (step-04). Once a user is successfully assigned a single subcarrier, it becomes an active user (step-07 and step-08). Then, the basestation stops join-in-testing for this user and starts searching bandwidth for other users (step-09). The sequence of testing users is based on our allocation-scheduling scheme, which will be described in Section 3.5 with more details. The users which cannot be assigned with any subcarrier are not allowed to transmit data during a block-interval. But, they will apply for data transmission when the next subchannel-allocation procedure is performed. We define these users within a block-interval as *inactive users*. It is clear that the index set of inactive users is  $\Omega \setminus \Phi$ . As a result, every active (inactive) user which is accepted (rejected) to transmit data is assigned one (zero) subcarrier during the initial allocation algorithm, which can be expressed as

$$\alpha_k = \frac{U_k}{U_k^{\max}} = \begin{cases} 1, & k \in \Phi \quad (\text{active user, } U_k = 1) \\ 0, & k \in \Omega \setminus \Phi \quad (\text{inactive user, } U_k = 0). \end{cases} \quad (33)$$

### 3.4. Dynamic allocation algorithm

The initial allocation algorithm maximizes the number of active users, but it considers neither optimizing the proportional bandwidth fairness nor maximizing the system throughput, which are the goals of our dynamic allocation algorithm. The key idea of our dynamic allocation is to assign all the leftover bandwidth resources after the initial allocation to active users based on the proportional fairness criterion [see Eq. (29)] defined in Section 3.2. The pseudocode of the proposed dynamic allocation algorithm is shown in Fig. 4.

Let  $\Theta$  denote the index set of active users which cannot be assigned with extra bandwidth resources. At the beginning of

```

00. Procedure: Dynamic Allocation Algorithms;
01.  $\Theta := \emptyset;$  ! Set the initial value of  $\Theta$ 
    ! Dynamic subcarrier allocation ends when there is no user which can be assigned any extra bandwidth
02. while ( $\Theta \neq \Phi$ ) do ! The algorithm ends when  $\Theta = \Phi$ 
03. begin ! find the active user which has the minimum aggressive factor and can be assigned extra bandwidth
04.  $k := \arg \min_{k \in \Phi \setminus \Theta} \alpha_k;$  ! Computing  $\alpha_k$  using Eq. (28)
05. if ( $\alpha_k \neq 1$ ) then ! The user does not achieve its maximum bandwidth requirement
06. begin for  $u := 1$  to  $U$  do ! Sequentially test subcarriers
07. begin Joining-Test for the  $k$ th user at the  $u$ th subcarrier ! Using Eq. (30)
08. if Eq. (30) is satisfied then
09. begin  $U_k := U_k + 1;$  ! Assign bandwidth to the  $k$ th user
10. Update  $\alpha_k;$  ! Using Eq. (28)
11. break; ! Find the next user which has the minimum  $\alpha_k$ 
12. end;
13. end;
14. if Eq. (30) is not satisfied then  $\Theta := \Theta + \{k\};$  ! No leftover bandwidth for the  $k$ th user
15. end
16. else  $\Theta := \Theta + \{k\};$  ! User achieves the maximum bandwidth requirement
17. end

```

**Fig. 4** Pseudocode of dynamic subchannel-allocation algorithms

dynamic allocation algorithm,  $\Theta$  is initialized as an empty set  $\Theta = \emptyset$  (step-01). If the  $k$ th user belongs to  $\Theta$  ( $k \in \Theta \subseteq \Phi$ ), it implies that:

**Case I:** The  $k$ th user has achieved its maximum bandwidth capacity, i.e.,  $U_k = U_k^{\max}$ , or

**Case II:** There is no leftover bandwidth available for the  $k$ th user.

The dynamic allocation algorithm searches for the  $k$ th user ( $k \in \Phi$ ), which has the *minimum* aggressive factor  $\alpha_k$  defined by Eq. (28) and can be assigned with extra bandwidth ( $k \notin \Theta$ ). As shown in Fig. 4, the index of this user can be obtained by (step-04):

$$k = \arg \min_{k \in \Phi \setminus \Theta} \{\alpha_k\} \quad (34)$$

where  $\Phi$  is the index set of active users obtained by the initial allocation algorithm.

If the  $k$ th user does not achieve its maximum bandwidth capacity (step-05), i.e.,  $U_k < U_k^{\max}$ , the algorithm at the basestation attempts to assign an extra subcarrier to it based on SINR criterion given by Eq. (30) (step-07). The attempt can succeed or fail. If it succeeds (step-08), the basestation updates the aggressive factor  $\alpha_k$  of the  $k$ th user and selects the new user which has the minimum  $\alpha_k$  using Eq. (34) (step-11). If it fails (step-14), the basestation adds the  $k$ th user to  $\Theta$  (**Case II**).

If the  $k$ th user has achieved its maximum bandwidth capacity (step-16), i.e.,  $U_k = U_k^{\max}$ , the basestation also adds

it to  $\Theta$  (**Case I**). The dynamic allocation algorithm repeats until  $\Theta = \Phi$ .

### 3.5. Delay-fairness-driven scheduling

Re-examining the subchannel-allocation algorithm described above, one can observe that when the initial allocation is applied, the user which is tested earlier receives the higher priority to become the active user than the later tested ones. This is due to the fact that SINR criterion is easier to be satisfied by the earlier tested users. If we employ a Round-Robin-based scheme testing all  $K$  users, the later attempted users are more likely to be rejected for transmitting data in each subchannel-allocation. As the result, their scheduling delays will be longer than the earlier attempted users. This problem gets more serious when the number  $K$  of users becomes larger. In the worst case, some users even cannot transmit any data at all.

To solve this problem, we propose the delay-fairness driven scheduling scheme to minimize the upper-bound of scheduling delay. Specifically, the scheduling is applied periodically to adjust the sequence of mobile users' attempts. During each interval of scheduling, the scheduling delays of all  $K$  users are recorded accumulatively and sorted. The complexity of the scheduling varies, depending on different scheduling algorithms used. The user which experiences larger delay will be tested earlier in initial allocations during the next interval of scheduling. Therefore, the scheduling delays will be fairly shared by all users, and thus this scheduling is called *delay-fairness driven* scheduling. Note the idea of

the proposed delay-fairness driven scheduling is similar to the earliest deadline first (EDF) [20] for the real-time services.

### 3.6. The tradeoff between fairness and throughput

Clearly, when the total number of users becomes larger, the system throughput will increase, but the bandwidth fairness will decrease due to the larger number of the inactive users. However, if we want to keep the proportional bandwidth fairness high, it will limit the total number of users assigned to the system. Then, we cannot achieve the maximum system throughput. This implies that there is an optimal tradeoff between fairness and throughput, which is specified by the optimal number of users  $K^*$  for the given system. To quantify this tradeoff, we define the *differential-constrained throughput*  $\theta$  by

$$\theta \triangleq \phi \bar{R} \tag{35}$$

where  $\bar{R}$  denotes the average system throughput. Because the fairness is a monotonic-decreasing, while the throughput is a monotonic-increasing, function of  $K$ , the product of  $\phi$  and  $\bar{R}$  has the maximum value, which is the optimal tradeoff between fairness and throughput. Using the parameter  $\theta$ , we can determine the optimal total number of users, which represents the best system traffic load, for the given SNR-threshold,  $\gamma$ , and the number of antennas as follows:

$$K^* = \arg \max_K \{\theta\} \tag{36}$$

### 3.7. Implementation issues

Our proposed subchannel allocation algorithms are executed by the basestation each time when it obtains the new channel state information (CSI) of the uplink channel. Then, the allocation decision is fed back to the each mobile user. It is clear that the feedback will introduce certain amount of overhead. However, such overhead can be relatively small, especially when the channel changes slowly (e.g., in an indoor slow-mobility environment), and the channel allocation can be done once every many symbol intervals [12]. Examining our proposed subchannel allocation algorithms, the computational complexity is mainly concentrated on the joining-test, i.e., Eqs. (31) and (32). Clearly, the complexity of this part is  $O(MN^2K)$ . However, in real wireless networks, the number  $N$  of transmit antennas at the mobile users is usually small due to the limited power supply and the cost constraint of handsets. Furthermore, the major algorithms are only executed at the basestation, which is much more powerful than the mobile users in handling more complex algorithms. Thus,

the proposed algorithms still remain practical for real wireless networks.

## 4. Simulation and numerical evaluations

We investigate the performance of the proposed system infrastructure and subchannel-allocation algorithms by simulations. In the simulations, the chip duration is set to  $T_c = 7.8125 \mu s$  and spreading gain  $G$  set to 16. Thus, the bit duration is  $T_b = T_c G = 125 \mu s$  and bit rate per subcarrier is 8 Kbps. We set the total number  $U$  of subcarriers equal to  $U = 8$ . Therefore, the total bandwidth of the system is  $U/T_c = 1.024$  MHz. The packet size is set to  $L = 8000$  bits and the SNR per bit  $\gamma_b$  is set to 10 dB at the basestation. The subchannel-allocation-scheduling is executed every 10 ms. The spreading codes are random signature sequences. The correlation factor between different asynchronous random spreading sequences is approximated as a Gaussian random variable with zero-mean and variance of  $1/(3G)$  [8]. Doppler frequency is set to 20 Hz for all users. Also, we omit the cyclic prefix (CP) for simplicity.

Figure 5 plots the simulated system throughput and numerically solved goodput [using Eq. (26)] versus SINR threshold  $\gamma$ . The plots of throughput are obtained through simulation experiments and the results of goodput are based on the numerical solutions for the analyses derived in Section 3.1. The total number  $K$  of mobile users is set to 15. As expected, the system throughput decreases when SINR threshold  $\gamma$  increases. However, the goodputs have the peak value when  $\gamma$  is around 10 dB, which represents the optimal SINR threshold  $\gamma_{opt} (\approx 10$  dB) for  $K = 15$ . We can also observe that the space-time infrastructure can significantly improve the system throughput and goodput, but the optimal SINR threshold  $\gamma_{opt} \approx 10$  dB virtually does not change regardless of what different antenna combinations are used.

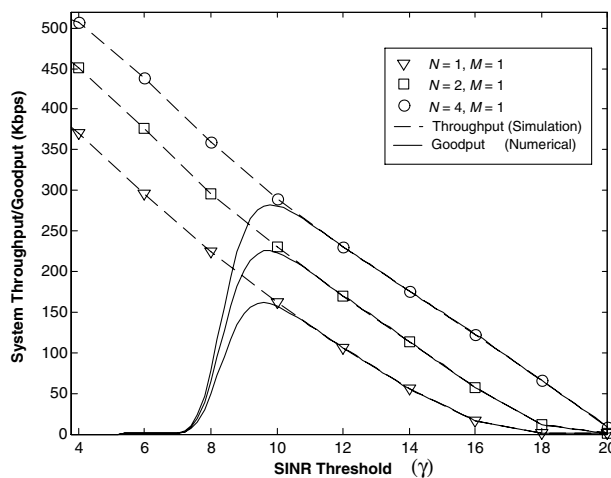
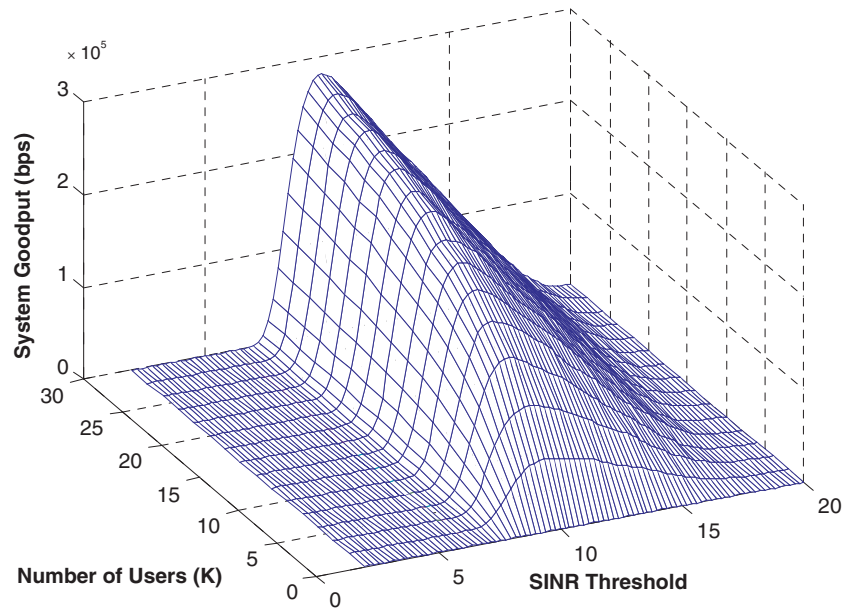


Fig. 5 System throughput  $\bar{R}$  and goodput  $\tilde{R}$  versus SINR threshold  $\gamma$

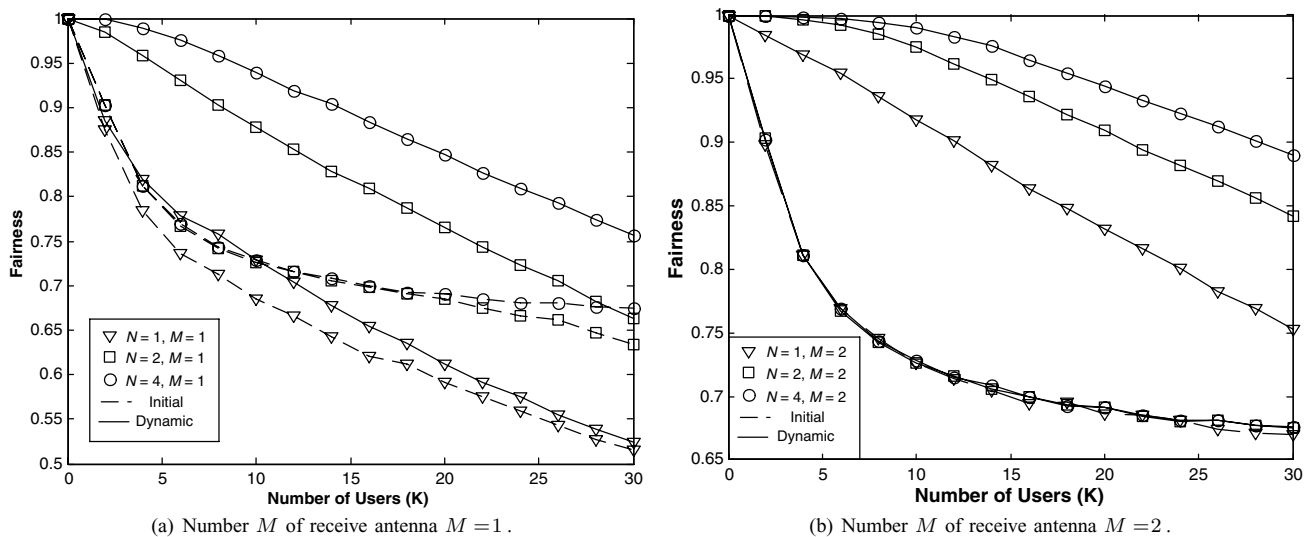
**Fig. 6** System goodput  $\tilde{R}$  versus SINR threshold  $\gamma$  and number  $K$  of users



To further investigate the influence of total number  $K$  of mobile users on the optimal SINR threshold  $\gamma_{opt}$ , Fig. 6 plots the numerically solved system goodput against two independent variables SINR threshold  $\gamma$  and the total number  $K$  of mobile users. The number of transmit antennas is set to  $N = 2$  and the number of receive antenna is  $M = 1$ . As shown in Fig. 6, the system goodput attains the maximum when  $\gamma \approx 10$  dB which is virtually independent of the total number  $K$  of mobile users. From Fig. 5 and Fig. 6, we can conclude that the optimal SINR threshold  $\gamma_{opt}$  can be pre-determined as a fixed value, regardless of what number of transmit antennas are used and what number of users are employed. Since the maximum goodput can be achieved, and the goodput approaches to the throughput when  $\gamma \approx 10$  dB,

thus, we set the SINR threshold  $\gamma$  to 10 dB and only investigate the system throughput ( $\approx$  goodput when  $\gamma = 10$  dB) in the following simulations for convenience.

Applying the initial and dynamic allocation algorithms, the simulated fairness and throughput versus the number  $K$  of mobile users with different numbers of transmit and receive antennas are presented in Figs. 7 and 8, respectively. The bandwidth capacities  $U_k^{max}$  are randomly chosen between 1 and  $U$ . We observe that the bandwidth fairness decreases, while the throughput increases, as  $K$  gets larger. This is expected since the larger number of users, the higher spectrum-efficiency of CDMA system, and thus the higher system throughput. On the other hand, the larger number of mobile users makes the bandwidth fairness more difficult to



**Fig. 7** Fairness  $\phi$  versus number  $K$  of users when the numbers of transmit and receive antennas vary









