Threshold-Based Opportunistic Scheduling for Ergodic Rate Guarantees in Wireless Networks

Yoora Kim, *Student Member, IEEE,* and Gang Uk Hwang, *Member, IEEE*

***Abstract*—In this paper, we propose an opportunistic downlink scheduling scheme that exploits multiuser diversity in a wireless network with threshold-based limited feedback. We assume that each user has its own ergodic rate requirement. The design objective of our scheme is to determine the values of thresholds with which heterogeneous ergodic rate requirements of all users are satisfied. In our analysis, we present a formula to check the feasibility of given ergodic rate requirements, and then obtain the feasible thresholds that realize them. We also obtain the optimal thresholds that maximize the ergodic sum-rate of the network while guaranteeing the ergodic rate requirements. Through numerical studies and simulations, we show the usefulness of our scheme and analysis.**

***Index Terms*—Ergodic rate guarantee, limited feedback, multiuser diversity, opportunistic scheduling, threshold.**

I. INTRODUCTION

IN wireless networks, to efficiently utilize the radio spectrum, opportunistic scheduling schemes exploit multiuser diversity by e.g., selecting only one user with the best channel condition at each time. Under this strategy, the total information-theoretic capacity of a wireless network can be maximized [1], [2].

Opportunistic scheduling schemes necessitate the base station (BS) to know the channel qualities of the mobile stations (MSs), which are estimated at the MSs and fed back to the BS. Hence, as the number of MSs increases, the feedback load becomes significant and yields the signaling overhead. Moreover, the power expended for feedback transmission by the non-scheduled MSs gets wasted [3]. To solve these problems, threshold-based feedback reduction algorithms have been studied extensively (see [4], [5] and the references therein). The main purpose of them is to reduce the number of MSs transmitting feedback while conserving the network performance. Consequently, the signaling overhead from feedback and the power consumption of MSs can be reduced. Motivated on this fact, we propose an opportunistic downlink scheduling scheme that exploits multiuser diversity with threshold-based limited feedback.

In this paper, we address the issue of guaranteeing ergodic rate requirements of all MSs by opportunistic scheduling with threshold-based limited feedback. The design objective of our scheduling scheme is to determine the values of thresholds with which ergodic rate requirements are satisfied for all MSs. By adjusting the values of thresholds, our scheme can support various ergodic rate gains, which shows an advantage of flexibility over the scheduling schemes, e.g., proportional fair scheduling (PFS), round-robin scheduling, that provide fixed ergodic rate gains.

The problem of guaranteeing ergodic rates was also addressed in [6], [7], where utility-based opportunistic scheduling schemes are presented. There are two main differences between our proposed scheme and those in [6], [7]. First, the utility functions used for scheduling decisions in [6], [7] are based on the full feedback information from *all* the MSs, while our scheme is based on the limited feedback information from a subset of the MSs. Second, our scheduling scheme uses thresholds as network parameter, but the scheduling schemes in [6], [7] use different network parameters such as token counter.

The remainder of this paper is organized as follows. In Section II, we propose our scheduling scheme and formulate the problems considered in this paper. We solve the problems for a heterogenous two-user case and a general N-user case in Sections III and IV, respectively. We provide numerical studies and simulations in Section V, and give conclusions in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider downlink transmission in a time-slotted wireless network consisting of a BS and N MSs, denoted by $\mathrm{MS}_n\,(n = 1, \ldots, N)$. In each time slot, a scheduler selects one MS, and the BS transmits data to the selected MS over a fading channel using a constant transmit power. In this paper, we assume a Rayleigh block fading channel, where the instantaneous received signal-to-noise ratio (SNR) $\gamma_n(t)$ of MS_n remains constant over time slot t but varies between time slots with the average received SNR $\overline{\gamma}_n$.

A. Proposed Scheduling Scheme

In our scheduling scheme, the BS sets the *a priori* threshold γ_{th_n} for MS_n. At time slot t, MS_n is allowed to feed back its instantaneous received SNR to the BS only when $\gamma_n(t) \geq \gamma_{th_n}$. We call the MS, who feeds back, the *feedback MS*. Among the feedback MS(s) at time slot t, the scheduler selects only one MS with the best instantaneous received SNR. If there is no feedback MS at time slot t, data are not transmitted at time slot t, in which case we declare a *scheduling outage*. In Sections III and IV, we show that the event of a scheduling outage can not happen under our

Manuscript received February 11, 2008; revised June 18, 2008 and September 2, 2008; accepted September 5, 2008. The associate editor coordinating the review of this paper and approving it for publication was V. Leung.

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R01-2007-000-20053-0).

The authors are with the Department of Mathematical Sciences and Telecommunication Engineering Program, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea (e-mail: jd_salinger@kaist.ac.kr; guhwang@kaist.edu).

Digital Object Identifier 10.1109/T-WC.2009.080191

scheduling scheme when the optimal thresholds are used. Our scheduling scheme is identical to the MaxSNR scheme if $\gamma_{th_1} = \ldots = \gamma_{th_N} = 0$.

B. Problem Formulation

In this paper, we focus on the ergodic rate performances of both the MSs and the network under our proposed scheduling scheme. Since we consider the *ergodic* rate, we drop the time index t in $\gamma_n(t)$ and use γ_n from now on. Under our proposed scheduling scheme, the ergodic rate of MS_n, denoted by $\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N})$ [b/s/Hz], can be expressed as follows:

$$\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N}) = \int_{\gamma_{th_n}}^{\infty} \log_2(1+x) \cdot \Pr\left(\max_k \{\gamma_k \cdot I(\gamma_k \geq \gamma_{th_k})\} = x \middle| \gamma_n = x \right) f_{\gamma_n}(x) dx \quad (1)$$

where $I(\cdot)$ denotes the indicator function, and $f_{\gamma_n}(x) := e^{-x/\overline{\gamma}_n}/\overline{\gamma}_n$ denotes the probability density function of the received SNR γ_n under the Rayleigh fading model. Here, the first term is the Shannon capacity per unit bandwidth with the received SNR $\gamma_n = x$. The second term represents the conditional probability that MS_n is selected by the scheduler given that its received SNR γ_n is x. The ergodic rate of MS_n is obtained by averaging the Shannon capacity per unit bandwidth over the distribution of the received SNR γ_n multiplied by the probability of being scheduled.

Suppose that MS_n has the ergodic rate requirement T_n [b/s/Hz]. In our model, it is assumed that the ergodic rate requirements T_n can be different from MS to MS. To clarify the problems considered in this paper, we introduce the following two definitions.

Definition 1: (Feasibility) A set of ergodic rate requirements $\{T_n\}_{n=1}^N$ is called *feasible* if there exists a corresponding set of thresholds that satisfy $\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N}) = T_n$ for all $1 \leq n \leq N$. Such thresholds $\{\gamma_{th_n}\}_{n=1}^N$ are called the *feasible thresholds*.

To emphasize the feasibility, we denote the feasible ergodic rate requirements and the corresponding feasible thresholds as $\{\tilde{T}_n\}_{n=1}^N$ and $\{\tilde{\gamma}_{th_n}\}_{n=1}^N$, respectively. Definition 1 says that if ergodic rate requirements $\{\tilde{T}_n\}_{n=1}^N$ are feasible, then the ergodic rate requirements of all MSs are guaranteed by choosing feasible thresholds $\{\tilde{\gamma}_{th_n}\}_{n=1}^N$. The region in $\mathbb{R}^N$ formed by $\{\tilde{T}_n\}_{n=1}^N$ is called the *feasible region* and denoted by $\mathbb{R}_{fsb}^N$. The first problem considered in this paper is to obtain the feasible region $\mathbb{R}_{fsb}^N$, hence we can easily check the feasibility of given ergodic rate requirements.

Definition 2: (Optimality) For given feasible ergodic rate requirements $\{\tilde{T}_n\}_{n=1}^N$ in $\mathbb{R}_{fsb}^N$, the thresholds, denoted by $\{\hat{\gamma}_{th_n}\}_{n=1}^N$, are called the *optimal thresholds* if

$$\begin{aligned} &\overline{C}_n(\gamma_{th_1}^*, \ldots, \gamma_{th_N}^*) \geq \tilde{T}_n, \qquad n = 1, \ldots, N, \\ &\overline{C}_{sum}(\hat{\gamma}_{th_1}, \ldots, \hat{\gamma}_{th_N}) = \max_{\{\gamma_{th_1}^*, \ldots, \gamma_{th_N}^*\}} \overline{C}_{sum}(\gamma_{th_1}^*, \ldots, \gamma_{th_N}^*), \end{aligned} \quad (2)$$

where $\overline{C}_{sum}(\cdot) := \sum_{n=1}^N \overline{C}_n(\cdot)$ denotes the ergodic sum-rate of the network.

To emphasize the optimality, for each feasible ergodic rate requirements $\{\tilde{T}_n\}_{n=1}^N$, we denote the ergodic rates obtained by using the corresponding optimal thresholds as $\{\hat{T}_n\}_{n=1}^N$ and call them the *optimal ergodic rates*. Obviously, we have $\hat{T}_n \geq \tilde{T}_n$ for all $1 \leq n \leq N$. Definition 2 says that, for given feasible ergodic rate requirements, the optimal thresholds can maximize the ergodic sum-rate of the network while guaranteeing the ergodic rate requirements of all MSs. The second problem considered in this paper is to obtain the optimal thresholds for given feasible ergodic rate requirements.

III. Heterogeneous Two-user Case

We start our analysis with a simple two-user model ($N = 2$). In this section, we assume that each MS is subject to independent but not necessarily identically distributed Rayleigh fading. Under this assumption, the ergodic rate expression in (1) when $n = 1$ can be rewritten as follows:

$$\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) = \int_{\gamma_{th_1}}^{\infty} \log_2(1+x)\big\{\Pr(\gamma_2 < \gamma_{th_2}) + \Pr(\gamma_2 \geq \gamma_{th_2}, \gamma_2 < x)\big\} f_{\gamma_1}(x) dx. \quad (3)$$

To compute the integral on the right-hand side of (3), we define $H(\overline{\gamma}, \gamma_{th}) := \int_{\gamma_{th}}^{\infty} \log_2(1+x)\, f_{\gamma_1}(x) dx$. Then, using a similar derivation as in [8, Appendix B], we have $H(\overline{\gamma}, \gamma_{th}) = \log_2(1+\gamma_{th}) e^{-\gamma_{th}/\overline{\gamma}} + e^{1/\overline{\gamma}} E_1\left((1+\gamma_{th})/\overline{\gamma}\right)/\ln 2$, where $E_1(x) := \int_x^{\infty} e^{-y}/y\, dy$ is the exponential integral of order 1. By using the function $H(\cdot,\cdot)$, the ergodic rate of MS_1 can be written as follows:

$$\begin{aligned} &\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) \\ &= H(\overline{\gamma}_1, \gamma_{th_1}) - \frac{\overline{\gamma}_2}{\overline{\gamma}_1 + \overline{\gamma}_2} \cdot H\left(\frac{\overline{\gamma}_1 \overline{\gamma}_2}{\overline{\gamma}_1 + \overline{\gamma}_2}, \max_i \gamma_{th_i}\right) \\ &\quad + e^{-\gamma_{th_2}/\overline{\gamma}_2} \cdot \left[H\left(\overline{\gamma}_1, \max_i \gamma_{th_i}\right) - H(\overline{\gamma}_1, \gamma_{th_1})\right]. \end{aligned} \quad (4)$$

Similarly, we can derive the ergodic rate of MS_2 as follows:

$$\begin{aligned} &\overline{C}_2(\gamma_{th_1}, \gamma_{th_2}) \\ &= H(\overline{\gamma}_2, \gamma_{th_2}) - \frac{\overline{\gamma}_1}{\overline{\gamma}_1 + \overline{\gamma}_2} \cdot H\left(\frac{\overline{\gamma}_1 \overline{\gamma}_2}{\overline{\gamma}_1 + \overline{\gamma}_2}, \max_i \gamma_{th_i}\right) \\ &\quad + e^{-\gamma_{th_1}/\overline{\gamma}_1} \cdot \left[H\left(\overline{\gamma}_2, \max_i \gamma_{th_i}\right) - H(\overline{\gamma}_2, \gamma_{th_2})\right]. \end{aligned} \quad (5)$$

Note that, when $\gamma_{th_1} \geq \gamma_{th_2}$, the ergodic rate of MS_1 in (4) is reduced to $\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) = H(\overline{\gamma}_1, \gamma_{th_1}) - \overline{\gamma}_2/(\overline{\gamma}_1 + \overline{\gamma}_2) \cdot H\left(\overline{\gamma}_1 \overline{\gamma}_2/(\overline{\gamma}_1 + \overline{\gamma}_2), \gamma_{th_1}\right)$, and depends only on the value of its threshold γ_{th_1} for the fixed $\overline{\gamma}_1$ and $\overline{\gamma}_2$. Similarly, the ergodic rate of MS_2 in (5) depends only on the value of its threshold γ_{th_2} when $\gamma_{th_1} \leq \gamma_{th_2}$. Hence, we use $\overline{C}_n(\gamma_{th_n})$ to denote $\overline{C}_n(\gamma_{th_1}, \gamma_{th_2})$ when $\arg\max_i \gamma_{th_i} = n$. As mentioned before, the ergodic rate gain of the MaxSNR scheme can be *feasible* under our proposed scheduling scheme by choosing the thresholds $\{\gamma_{th_1}, \gamma_{th_2}\} = \{0, 0\}$, and is given by $\overline{C}_{n,\text{MaxSNR}} = \overline{C}_n(0,0) = H(\overline{\gamma}_n, 0) - \overline{\gamma}_1 \overline{\gamma}_2/(\overline{\gamma}_n(\overline{\gamma}_1 + \overline{\gamma}_2)) \cdot H\left(\overline{\gamma}_1 \overline{\gamma}_2/(\overline{\gamma}_1 + \overline{\gamma}_2), 0\right)$ $(n = 1, 2)$.

By using $\overline{C}_n(\gamma_{th_1}, \gamma_{th_2})$ $(n = 1, 2)$ in (4) and (5), we can obtain the feasible region $\mathbb{R}_{fsb}^2$ as follows.

Theorem 1: The feasible region $\mathbb{R}_{fsb}^2$ is given by

$$\mathbb{R}_{fsb}^2 = \{(x, y) \in \mathbb{R}^2 \,|\, 0 \leq x \leq \overline{C}_{1,bd}, 0 \leq y \leq f(x)\}$$

where $\overline{C}_{1,bd} := \lim_{\gamma_{th_2}\to\infty} \overline{C}_1(0,\gamma_{th_2}) = H(\overline{\gamma}_1, 0)$ represents the upper bound of $\overline{C}_1(\cdot,\cdot)$, and

$$f(x) := \max_{\overline{C}_1(\gamma_{th_1},\gamma_{th_2})=x} \overline{C}_2(\gamma_{th_1},\gamma_{th_2}) = \begin{cases} \overline{C}_2(\gamma^*_{th_1,x}, 0), & 0 \le x \le \overline{C}_{1,\text{MaxSNR}}, \\ \overline{C}_2(0, \gamma^*_{th_2,x}), & \overline{C}_{1,\text{MaxSNR}} < x \le \overline{C}_{1,bd}. \end{cases}$$

Here, $\gamma^*_{th_1,x}$ and $\gamma^*_{th_2,x}$ are uniquely determined from $\overline{C}_1(\gamma^*_{th_1,x},0) = x$ and $\overline{C}_1(0,\gamma^*_{th_2,x}) = x$.

Proof: See Appendix A. ■

In Fig. 1, as an example, we show the feasible region $\mathbb{R}^2_{fsb}$ for $\overline{\gamma}_1 = 15$ dB and $\overline{\gamma}_2 = 20$ dB by using Theorem 1. The result in Fig. 1 is discussed in more detail in Section VI. A.

Our proposed scheduling scheme can guarantee *any* ergodic rate requirements that are located in the feasible region by properly choosing the corresponding feasible thresholds. Theorem 2 gives a formula to obtain the feasible thresholds.

Theorem 2: For given feasible ergodic rate requirements $\{\tilde{T}_1, \tilde{T}_2\}$ in $\mathbb{R}^2_{fsb}$, the corresponding feasible thresholds $\{\tilde{\gamma}_{th_1}, \tilde{\gamma}_{th_2}\}$ are uniquely determined from $\overline{C}_n(\tilde{\gamma}_{th_n}) = \tilde{T}_n$ and $\overline{C}_m(\tilde{\gamma}_{th_1}, \tilde{\gamma}_{th_2}) = \tilde{T}_m$, where $\arg\max_i \tilde{\gamma}_{th_i} = n$ and $\arg\min_i \tilde{\gamma}_{th_i} = m$. Further, they satisfy $\max\{\tilde{\gamma}_{th_1}, \tilde{\gamma}_{th_2}\} = \tilde{\gamma}_{th_1}$ if $\tilde{T}_1 \le \overline{C}_{1,\text{MaxSNR}}$ and $\tilde{T}_2 \ge \overline{C}_2(\gamma^*_{th})$, where γ^*_{th} is uniquely determined from $\overline{C}_1(\gamma^*_{th}) = \tilde{T}_1$. Otherwise, they satisfy $\max\{\tilde{\gamma}_{th_1}, \tilde{\gamma}_{th_2}\} = \tilde{\gamma}_{th_2}$.

Proof: From the proof of Theorem 1, our theorem immediately follows. ■

For given feasible ergodic rate requirements, the use of optimal thresholds can maximize the ergodic sum-rate of the network while guaranteeing the ergodic rate requirements of all MSs. Theorem 3 gives a formula to obtain the optimal thresholds.

Theorem 3: For given feasible ergodic rate requirements $\{\tilde{T}_1, \tilde{T}_2\}$ in $\mathbb{R}^2_{fsb}$, the optimal thresholds $\{\hat{\gamma}_{th_1}, \hat{\gamma}_{th_2}\}$ are given by

$$\{\hat{\gamma}_{th_1}, \hat{\gamma}_{th_2}\} = \begin{cases} \{0,0\}, & \text{if } \tilde{T}_1 \le \overline{C}_{1,\text{MaxSNR}} \text{ and } \tilde{T}_2 \le \overline{C}_{2,\text{MaxSNR}}, \\ \{\gamma^*_{th_1}, 0\}, & \text{if } \tilde{T}_1 \le \overline{C}_{1,\text{MaxSNR}} \text{ and } \tilde{T}_2 > \overline{C}_{2,\text{MaxSNR}}, \\ \{0, \gamma^*_{th_2}\}, & \text{if } \tilde{T}_1 > \overline{C}_{1,\text{MaxSNR}} \text{ and } \tilde{T}_2 \le \overline{C}_{2,\text{MaxSNR}}, \end{cases}$$

where $\gamma^*_{th_1}$ and $\gamma^*_{th_2}$ are uniquely determined from $\overline{C}_2(\gamma^*_{th_1}, 0) = \tilde{T}_2$ and $\overline{C}_1(0, \gamma^*_{th_2}) = \tilde{T}_1$.

Proof: See Appendix B. ■

The result in Theorem 3 implies that the event of a scheduling outage can not happen under our scheduling scheme when the optimal thresholds are used.

IV. General N-user Case

In this section, we assume a general N-user model ($N \ge 2$) where each MS is subject to independent and identically distributed Rayleigh fading, i.e., $\overline{\gamma}_1 = \overline{\gamma}_2 = \ldots = \overline{\gamma}_N$ ($:= \overline{\gamma}$). Without loss of generality, we assume that $T_1 \le T_2 \le \ldots \le T_N$ or, equivalently, $\gamma_{th_1} \ge \gamma_{th_2} \ge \ldots \ge \gamma_{th_N}$. Under these assumptions, the ergodic rate of MS_n in (1) can be rewritten as follows:

$$\begin{aligned} &\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N}) \\ &= \int_{\gamma_{th_1}}^{\infty} \log_2(1+x) \prod_{l=1, l\neq n}^{N} \Pr(\gamma_l < x) f_{\gamma_n}(x)dx \\ &\quad + \sum_{i=2}^{n} \prod_{k=1}^{i-1} \Pr(\gamma_k < \gamma_{th_k}) \int_{\gamma_{th_i}}^{\gamma_{th_{i-1}}} \log_2(1+x) \\ &\quad \cdot \prod_{l=i, l\neq n}^{N} \Pr(\gamma_l < x) f_{\gamma_n}(x)dx. \end{aligned} \tag{6}$$

Let

$$G_i(\gamma_{th_1}, \ldots, \gamma_{th_i}) := \prod_{k=1}^{i-1} \Pr(\gamma < \gamma_{th_k}) \int_{\gamma_{th_i}}^{\gamma_{th_{i-1}}} \log_2(1+x) \cdot [\Pr(\gamma < x)]^{N-i} f_\gamma(x)dx$$

where γ denotes the generic random variable for γ_n. Then, we have

$$G_i(\gamma_{th_1}, \ldots, \gamma_{th_i}) = \prod_{k=1}^{i-1} [1 - e^{-\gamma_{th_k}/\overline{\gamma}}] \sum_{j=0}^{N-i} \binom{N-i}{j} \cdot \frac{(-1)^j}{j+1} \left[H\left(\overline{\gamma}/(j+1), \gamma_{th_i}\right) - H\left(\overline{\gamma}/(j+1), \gamma_{th_{i-1}}\right) \right].$$

By using the function $G_i(\cdot)$, the ergodic rate of MS_n in (6) can be written in a closed-form as follows:

$$\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N}) = \sum_{i=1}^{n} G_i(\gamma_{th_1}, \ldots, \gamma_{th_i}). \tag{7}$$

Note from (7) that the ergodic rate of MS_n is determined only from the values of thresholds $\{\gamma_{th_i}\}_{i=1}^n$ for the fixed $\overline{\gamma}$. Hence, instead of using $\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_N})$, we use $\overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_n})$ to denote the ergodic rate of MS_n. By using (7), we can obtain the feasible region $\mathbb{R}^N_{fsb}$ as in the following theorem.

Theorem 4: The feasible region $\mathbb{R}^N_{fsb}$ is given by

$$\begin{aligned} \mathbb{R}^N_{fsb} = \{ &(x_1, \ldots, x_N) \in \mathbb{R}^N \,|\, 0 \le x_1 \le \overline{C}_1(0), \\ &x_n \le x_{n+1} \le \overline{C}_{n+1}(\gamma^*_{th_1,x_1}, \ldots, \gamma^*_{th_n,x_n}, 0), \\ &n = 1, \ldots, N-1 \} \end{aligned}$$

where $\{\gamma^*_{th_n,x_n}\}_{n=1}^{N-1}$ are uniquely determined by initially obtaining $\gamma^*_{th_1,x_1}$ from $\overline{C}_1(\gamma^*_{th_1,x_1}) = x_1$ and then iteratively solving the equation $\overline{C}_n(\gamma^*_{th_1,x_1}, \ldots, \gamma^*_{th_n,x_n}) = x_n$ for $n = 2, \ldots, N-1$.

Using a similar approach as in the proof of Theorem 1, we can prove Theorem 4 (for the detailed derivation, see [9]). In Fig. 3, as an example, we show the feasible region $\mathbb{R}^3_{fsb}$ for $\overline{\gamma} = 15$ dB by using Theorem 4. The result in Fig. 3 is discussed in more detail in Section VI. B.

According to Definition 1, the ergodic rate requirements of MSs are in the feasible region $\mathbb{R}^N_{fsb}$ if and only if there exist the corresponding feasible thresholds $\{\tilde{\gamma}_{th_n}\}_{n=1}^N$. From the proof of Theorem 4, we have Theorem 5 which summarizes the existence and uniqueness of the corresponding feasible thresholds $\{\tilde{\gamma}_{th_n}\}_{n=1}^N$ for given feasible ergodic rate requirements $\{\tilde{T}_n\}_{n=1}^N$.

Theorem 5: For given feasible ergodic rate requirements $\{\tilde{T}_n\}_{n=1}^N$, the corresponding feasible thresholds $\{\tilde{\gamma}_{th_n}\}_{n=1}^N$ are uniquely determined from

$$\overline{C}_n(\tilde{\gamma}_{th_1}, \ldots, \tilde{\gamma}_{th_n}) = \tilde{T}_n, \qquad 1 \le n \le N. \tag{8}$$

Based on Theorem 5, we can simultaneously investigate the existence of the feasible thresholds and calculate the feasible thresholds successively from $\tilde{\gamma}_{th_1}$ to $\tilde{\gamma}_{th_N}$ by initially obtaining $\tilde{\gamma}_{th_1}$ from $\overline{C}_1(\tilde{\gamma}_{th_1}) = T_1$ and then iteratively solving the equation in (8) for $n = 2, \ldots, N$.

We now focus on the optimality problem given in (2). To solve the optimality problem, we restrict the domain of the function $\overline{C}_{sum}(\cdot)$ to the set $D := \{(\gamma_{th_1}, \ldots, \gamma_{th_N}) \in \mathbb{R}^N \mid \overline{C}_n(\gamma_{th_1}, \ldots, \gamma_{th_n}) \ge \tilde{T}_n, n = 1, \ldots, N\}$. It is easily checked that the domain D is closed and bounded. Since the function $\overline{C}_{sum}(\cdot)$ is differentiable, the maximal point exists in the domain D by the Maximum-Minimum Theorem [10]. Therefore, we can apply the Lagrange multiplier method [10] to find the maximal point, i.e., optimal thresholds, with the help of the following theorem.

Theorem 6: For given feasible ergodic rates $\{\tilde{T}_n\}_{n=1}^N$ in $\mathbb{R}_{fsb}^N$, suppose $\tilde{T}_N > \overline{C}_N(0, \ldots, 0)$. Then, the corresponding optimal thresholds $\{\hat{\gamma}_{th_n}\}_{n=1}^N$ exist in the boundary of domain D satisfying

$$\overline{C}_N(\hat{\gamma}_{th_1}, \ldots, \hat{\gamma}_{th_N}) = \tilde{T}_N \text{ and } \hat{\gamma}_{th_N} = 0. \tag{9}$$

If $\tilde{T}_N \le \overline{C}_N(0, \ldots, 0)$, then the optimal thresholds are given by $\{\hat{\gamma}_{th_1}, \ldots, \hat{\gamma}_{th_N}\} = \{0, \ldots, 0\}$.

Using a similar approach as in the proof of Theorem 3, we can prove Theorem 6 (for the detailed derivation, see [9]). The result in Theorem 6 implies that the event of a scheduling outage can not happen under our scheduling scheme when the optimal thresholds are used. From Theorem 6, when $\tilde{T}_N > \overline{C}_N(0, \ldots, 0)$, by the Lagrange multiplier method there exists a number $\lambda \in \mathbb{R}$ such that

$$\nabla\overline{C}_{sum}(\hat{\gamma}_{th_1}, \ldots, \hat{\gamma}_{th_{N-1}}, 0) = \lambda\nabla\overline{C}_N(\hat{\gamma}_{th_1}, \ldots, \hat{\gamma}_{th_{N-1}}, 0). \tag{10}$$

By solving the system of equations given in (9) and (10), we can obtain the optimal thresholds $\{\hat{\gamma}_{th_n}\}_{n=1}^{N-1}$.

Our scheduling scheme is applicable even in cases where the number of MSs changes over time. Suppose that the network consists of (initially) N MSs with feasible ergodic rate requirements. We first consider the case where an MS leaves the network after the completion of data transmission. Then, the feasible (resp. optimal) thresholds are adjusted for remaining $N-1$ MSs by using Theorem 5 (resp. 6). We next consider the case where an MS requests to enter the network. The incoming MS is accepted only if the ergodic rate requirements of all MSs are still feasible after the acceptance of the incoming MS. This is easily checked by using Theorem 4. If the incoming MS is accepted, then the feasible (resp. optimal) thresholds are adjusted for all $N+1$ MSs by using Theorem 5 (resp. 6).

V. SIMULATION RESULTS

A. Scenario 1: Heterogeneous Two-user Case

In this scenario, we consider a two-user case as in Section III, assuming $\overline{\gamma}_1 = 15$ dB and $\overline{\gamma}_2 = 20$ dB. The corresponding feasible region $\mathbb{R}_{fsb}^2$ is plotted in Fig. 1 by using Theorem 1. As shown in the figure, the shape of the feasible region depends on the average received SNR values and skews to the MS with higher average received SNR value. Suppose that the ergodic rate requirements of MS_1 and MS_2 are given by 3.0 and 1.0 b/s/Hz, respectively. Then, both MSs can achieve their target ergodic rates since they are in the feasible region. The corresponding feasible and optimal thresholds can be obtained from Theorems 2 and 3, respectively, and are given by $\{\tilde{\gamma}_{th_1}, \tilde{\gamma}_{th_2}\} = \{11.37, 23.24\}$ dB and $\{\hat{\gamma}_{th_1}, \hat{\gamma}_{th_2}\} = \{-\infty, 20.68\}$ dB. In Fig. 2, we show the resulting feasible and optimal ergodic rates obtained from analysis and simulation, where simulation results are based on the Zheng and Xiao's Rayleigh fading model [11]. As shown in the figure, both MSs can realize their target ergodic rates by using the feasible thresholds. Moreover, by using the optimal thresholds, MSs can achieve ergodic rates more than their target ergodic rates.

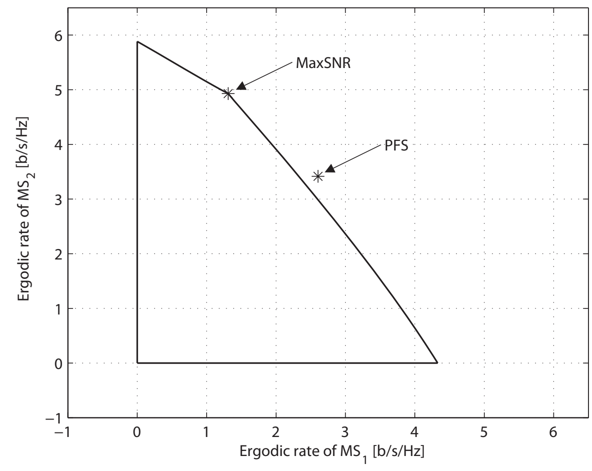

Fig. 1. Feasible region $\mathbb{R}_{fsb}^2$ for $\overline{\gamma}_1 = 15$ dB and $\overline{\gamma}_2 = 20$ dB. $*$'s are the ergodic rate gains of the MaxSNR scheme and the PFS scheme.

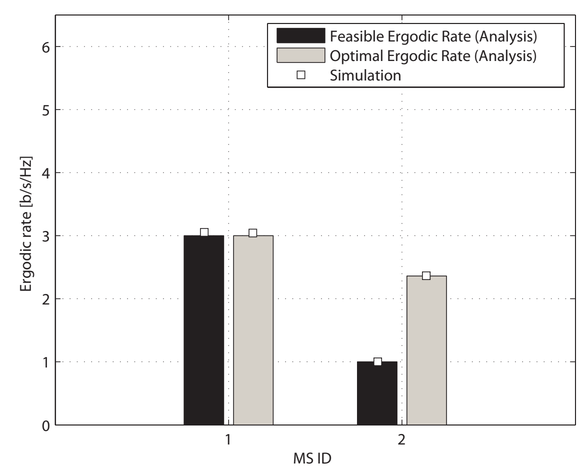

Fig. 2. Feasible and optimal ergodic rates for $\overline{\gamma}_1 = 15$ dB and $\overline{\gamma}_2 = 20$ dB.

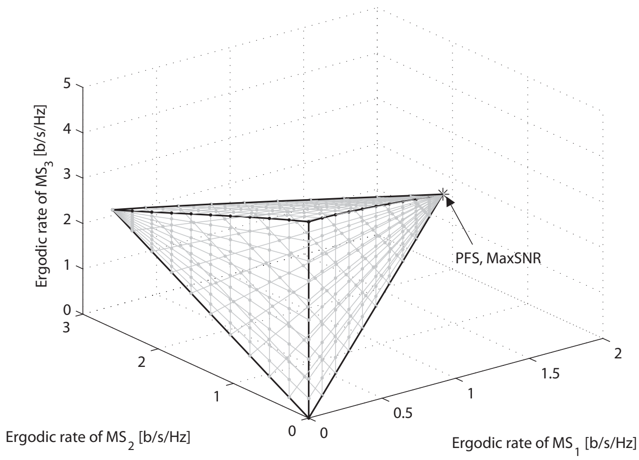

Fig. 3. Feasible region $\mathbb{R}^3_{fsb}$ for $\overline{\gamma}_1 = \overline{\gamma}_2 = \overline{\gamma}_3 = 15$ dB. $*$ is the ergodic rate gain of the MaxSNR scheme, and is also that of the PFS scheme.

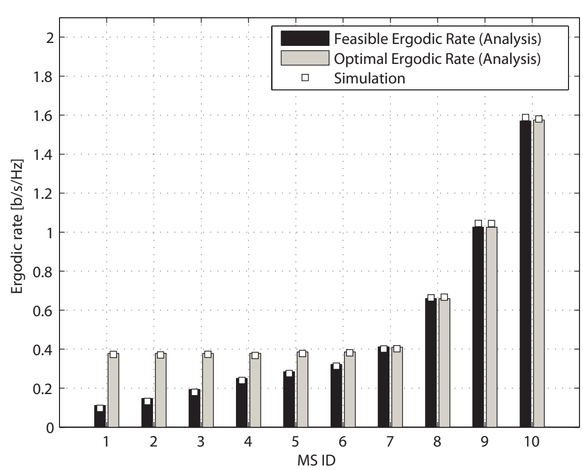

Fig. 4. Feasible and optimal ergodic rates for $\overline{\gamma}_n = 15$ dB ($n = 1, \ldots, 10$).

B. Scenario 2: General N-user Case

In this scenario, we consider a general N-user case as in Section IV, assuming $\overline{\gamma} = 15$ dB. For $N = 3$, the feasible region $\mathbb{R}^3_{fsb}$ is plotted in Fig. 3 by using Theorem 4. Suppose next that the network consists of 5 MSs with the ergodic rate requirements given by 0.5, 1.2, 2.5, 2.6 and 2.8 b/s/Hz, respectively. We can not plot the feasible region $\mathbb{R}^5_{fsb}$, but by using Theorem 5 we can logically check the feasibility of given ergodic rate requirements. In this case, we can obtain the feasible thresholds only for the MS_1 and MS_2, and accordingly the given ergodic rate requirements can not be feasible under our scheme.

Now suppose that the network consists of 10 MSs with the ergodic rate requirements given in Table I. In this case, the ergodic rates are feasible under our scheduling scheme, and we can obtain the feasible and optimal thresholds from Theorems 5 and 6, respectively, which are summarized in Table I. In Fig. 4, we show the resulting feasible and optimal ergodic rates obtained from analysis and simulation.

TABLE I
FEASIBLE AND OPTIMAL THRESHOLDS

MS ID	T_n	$\tilde{\gamma}_{th_n}$	$\hat{\gamma}_{th_n}$	MS ID	T_n	$\tilde{\gamma}_{th_n}$	$\hat{\gamma}_{th_n}$
1	0.11	21.14	19.08	6	0.32	19.54	19.03
2	0.15	20.79	19.08	7	0.41	19.03	18.86
3	0.19	20.41	19.08	8	0.66	17.78	17.24
4	0.25	20.00	19.08	9	1.03	16.02	14.62
5	0.28	19.78	19.03	10	1.57	13.01	$-\infty$

T_n [b/s/Hz]: the ergodic rate requirements
$\tilde{\gamma}_{th_n}$ [dB]: the feasible thresholds
$\hat{\gamma}_{th_n}$ [dB]: the optimal thresholds

VI. CONCLUSION

In this paper, we propose an opportunistic downlink scheduling scheme that can guarantee ergodic rate requirements of all MSs with threshold-based limited feedback. Analytic results are provided to check the feasibility of given ergodic rate requirements, and to obtain the corresponding optimal thresholds. Our numerical and simulation results show the usefulness of our proposed scheme and analysis.

APPENDIX A
PROOF OF THEOREM 1

In order to prove Theorem 1, we need the following lemmas.

Lemma 1: We have $\frac{\partial}{\partial \gamma_{th_m}} \overline{C}_n(\gamma_{th_1}, \gamma_{th_2}) < 0$ for $m = n$ and $\frac{\partial}{\partial \gamma_{th_m}} \overline{C}_n(\gamma_{th_1}, \gamma_{th_2}) > 0$ for $m \neq n$.

The proof of Lemma 1 is straightforward and omitted (for the detailed derivation, see [9]).

Lemma 2: Let for $\arg\max_i \gamma_{th_i} = n$,

$$\overline{C}_n(\gamma_{th_n}) := H(\overline{\gamma}_n, \gamma_{th_n}) - \frac{\overline{\gamma}_1 \overline{\gamma}_2}{\overline{\gamma}_n(\overline{\gamma}_1 + \overline{\gamma}_2)} H\left(\frac{\overline{\gamma}_1 \overline{\gamma}_2}{\overline{\gamma}_1 + \overline{\gamma}_2}, \gamma_{th_n}\right).$$

Then, $\overline{C}_n(\gamma_{th_n})$ is a strictly decreasing continuous function.

Proof: By the definition of H and Lemma 1 with $m = n$, Lemma 2 immediately follows. ■

We now prove Theorem 1. First consider $\gamma_{th_1} \geq \gamma_{th_2} \geq 0$. Then, $\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) = \overline{C}_1(\gamma_{th_1})$, and by Lemma 2, we have $0 = \lim_{\gamma_{th_1} \to \infty} \overline{C}_1(\gamma_{th_1}) \leq \overline{C}_1(\gamma_{th_1}) \leq \lim_{\gamma_{th_1} \to 0} \overline{C}_1(\gamma_{th_1}) = \overline{C}_{1,\mathrm{MaxSNR}}$. For each $x \in [0, \overline{C}_{1,\mathrm{MaxSNR}}]$, there exists the unique $\gamma^*_{th_1,x}$ such that $\overline{C}_1(\gamma^*_{th_1,x}) = x$ by Lemma 2. By Lemma 1, it follows that

$$\max_{\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) = x} \overline{C}_2(\gamma_{th_1}, \gamma_{th_2}) = \lim_{\gamma_{th_2} \to 0} \overline{C}_2(\gamma^*_{th_1,x}, \gamma_{th_2}) = \overline{C}_2(\gamma^*_{th_1,x}, 0),$$

and

$$\min_{\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) = x} \overline{C}_2(\gamma_{th_1}, \gamma_{th_2}) = \lim_{\gamma_{th_2} \to \gamma^*_{th_1,x}} \overline{C}_2(\gamma^*_{th_1,x}, \gamma_{th_2}) = \overline{C}_2(\gamma^*_{th_1,x}).$$

Therefore, for $0 \leq x \leq \overline{C}_{1,\mathrm{MaxSNR}}$, we obtain

$$\overline{C}_2(\gamma^*_{th_1,x}) \leq \overline{C}_2(\gamma^*_{th_1,x}, \gamma_{th_2}) \leq \overline{C}_2(\gamma^*_{th_1,x}, 0). \quad (11)$$

In addition, $\overline{C}_2(\gamma^*_{th_1,x}, 0) \geq \lim_{\gamma_{th_1} \to 0} \overline{C}_2(\gamma_{th_1}, 0) = \overline{C}_{2,\mathrm{MaxSNR}}$.

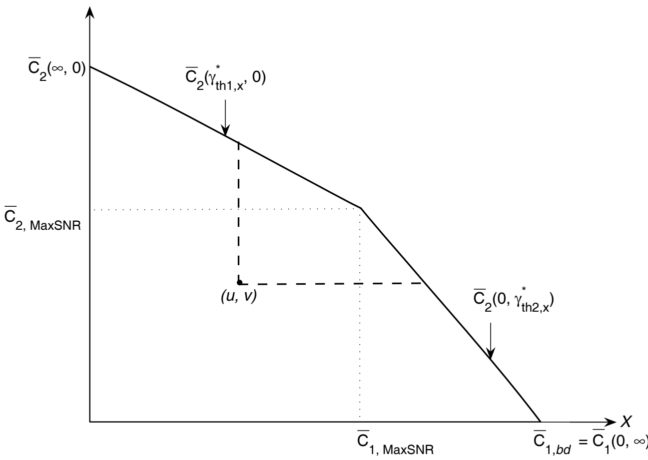

Fig. 5. Feasible region in the proof of Theorem 1.

Next consider $0 \leq \gamma_{th_1} < \gamma_{th_2}$. Then, $\overline{C}_2(\gamma_{th_1}, \gamma_{th_2}) = \overline{C}_2(\gamma_{th_2})$, and similarly as above we have $0 \leq \overline{C}_2(\gamma_{th_2}) \leq \overline{C}_{2,\mathrm{MaxSNR}}$. For each $y \in [0, \overline{C}_{2,\mathrm{MaxSNR}}]$, similarly as above there exists the unique $\gamma^*_{th_2,y}$ such that $\overline{C}_2(\gamma^*_{th_2,y}) = y$, and we obtain

$$\overline{C}_1(\gamma^*_{th_2,y}) \leq \overline{C}_1(\gamma_{th_1}, \gamma^*_{th_2,y}) \leq \overline{C}_1(0, \gamma^*_{th_2,y}). \quad (12)$$

In addition, $\overline{C}_{1,\mathrm{MaxSNR}} = \lim_{\gamma_{th_2} \to 0} \overline{C}_1(0, \gamma_{th_2}) \leq \overline{C}_1(0, \gamma^*_{th_2,y}) \leq \lim_{\gamma_{th_2} \to \infty} \overline{C}_1(0, \gamma_{th_2}) = \overline{C}_{1,bd}$.

For each point $(u, v) \in [0, \overline{C}_{1,\mathrm{MaxSNR}}] \times [0, \overline{C}_{2,\mathrm{MaxSNR}}]$ such that $\overline{C}_1(\gamma_{th}) = u$ and $\overline{C}_2(\gamma_{th}) = v$, we see from (11) and (12) that two sets $\{(u, s) \in \mathbb{R}^2 \,|\, v = \overline{C}_2(\gamma_{th}) \leq s \leq \overline{C}_2(\gamma_{th}, 0)\}$ and $\{(t, v) \in \mathbb{R}^2 \,|\, u = \overline{C}_1(\gamma_{th}) \leq t \leq \overline{C}_1(0, \gamma_{th})\}$ are included in the feasible region as shown in Fig. 5.

In summary, $(x, y) \in \mathbb{R}^2_{fsb}$ can be expressed as follows. If $0 \leq x \leq \overline{C}_{1,\mathrm{MaxSNR}}$, there exists the unique $\gamma^*_{th_1,x}$ such that $\overline{C}_1(\gamma^*_{th_1,x}) = x$, and we have $0 \leq y \leq \overline{C}_2(\gamma^*_{th_1,x}, 0)$. If $\overline{C}_{1,\mathrm{MaxSNR}} < x \leq \overline{C}_{1,bd}$, there exists the unique $\gamma^*_{th_2,x}$ such that $\overline{C}_1(0, \gamma^*_{th_2,x}) = x$, and we have $0 \leq y \leq \overline{C}_2(\gamma^*_{th_2,x})$.

Appendix B
Proof of Theorem 3

In order to prove Theorem 3, we need the following lemma.

Lemma 3: For $n = 1, 2$, we have $\frac{\partial}{\partial \gamma_{th_n}} \overline{C}_{sum}(\gamma_{th_1}, \gamma_{th_2}) < 0$. Hence, we obtain

$$\max_{\{\gamma_{th_1} \geq 0, \gamma_{th_2} \geq 0\}} \overline{C}_{sum}(\gamma_{th_1}, \gamma_{th_2}) = \overline{C}_{sum}(0, 0) = \overline{C}_{1,\mathrm{MaxSNR}} + \overline{C}_{2,\mathrm{MaxSNR}}.$$

The proof of Lemma 3 is straightforward and omitted (for the detailed derivation, see [9]).

Now we prove Theorem 3. For given feasible ergodic rate requirements $\{\tilde{T}_1, \tilde{T}_2\}$, we restrict the domain of $\overline{C}_{sum}(\gamma_{th_1}, \gamma_{th_2})$ as $D := \{(\gamma_{th_1}, \gamma_{th_2}) \in \mathbb{R}^2 \,|\, \overline{C}_n(\gamma_{th_1}, \gamma_{th_2}) \geq \tilde{T}_n,\, n = 1, 2\}$.

For $0 \leq \tilde{T}_1 \leq \overline{C}_{1,\mathrm{MaxSNR}}$ and $0 \leq \tilde{T}_2 \leq \overline{C}_{2,\mathrm{MaxSNR}}$, it is obvious that $\{0, 0\} \in D$. Hence, by Lemma 3, we have $\{\hat{\gamma}_{th_1}, \hat{\gamma}_{th_2}\} = \{0, 0\}$.

Now suppose that $0 \leq \tilde{T}_1 \leq \overline{C}_{1,\mathrm{MaxSNR}}$ and $\tilde{T}_2 > \overline{C}_{2,\mathrm{MaxSNR}}$. Then, for any $\{\gamma_{th_1}, \gamma_{th_2}\} \in D$, we have $\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) < \overline{C}_{1,\mathrm{MaxSNR}}$. Otherwise, we have $\overline{C}_{sum}(\gamma_{th_1}, \gamma_{th_2}) = \overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) + \overline{C}_2(\gamma_{th_1}, \gamma_{th_2}) > \overline{C}_{1,\mathrm{MaxSNR}} + \overline{C}_{2,\mathrm{MaxSNR}}$ ($\because$ $\overline{C}_{2,\mathrm{MaxSNR}} < \tilde{T}_2 \leq \overline{C}_2(\gamma_{th_1}, \gamma_{th_2})$), which is a contradiction by Lemma 3. Note that the maximum value of $\overline{C}_{sum}(\gamma_{th_1}, \gamma_{th_2})$ occurs when $\{\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}), \overline{C}_2(\gamma_{th_1}, \gamma_{th_2})\}$ is a boundary point of the feasible region by Lemma 3 or Theorem 1. Hence, in this case where $\overline{C}_1(\gamma_{th_1}, \gamma_{th_2}) < \overline{C}_{1,\mathrm{MaxSNR}}$, we have $\hat{\gamma}_{th_2} = 0$ by Theorem 1. Now we consider $\overline{C}_{sum}(\gamma_{th_1}, 0)$. By Lemma 3, the maximum value of $\overline{C}_{sum}(\gamma_{th_1}, 0)$ occurs when γ_{th_1} is the minimum value of the set $\{x \,|\, (x, 0) \in D\}$. That is, the optimal thresholds $\{\hat{\gamma}_{th_1}, \hat{\gamma}_{th_2}\}$ satisfy $\hat{\gamma}_{th_2} = 0$ and $\overline{C}_2(\hat{\gamma}_{th_1}, 0) = \tilde{T}_2$ by Theorem 1.

For the case where $\tilde{T}_1 > \overline{C}_{1,\mathrm{MaxSNR}}$ and $0 \leq \tilde{T}_2 \leq \overline{C}_{2,\mathrm{MaxSNR}}$, we can prove similarly as above and omit detailed derivations.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which significantly improved the presentation of this paper.

References

[1] R. Knopp and P. A. Humblet, “Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC*, Seattle, June 1995.

[2] D. N. Tse, “Optimal power allocation over parallel Gaussian broadcast channels," in *Proc. IEEE Int. Symp. on Inf. Theory*, Ulm, Germany, June 1997, pp. 27.

[3] H. Kim and Y. Han, “An opportunistic channel quality feedback scheme for proportional fair scheduling," *IEEE Commun. Lett.*, vol. 11, no. 6, pp. 501-503, June 2007.

[4] V. Hassel, M.-S. Alouini, G. E. Øien, and D. Gesbert, “Rate-optimal multiuser scheduling with reduced feedback load and analysis of delay effects," *EURASIP J. Wireless Commun. and Networking*, vol. 2006, Article ID 36424, 7 pages, 2006.

[5] V. Hassel, D. Gesbert, M.-S. Alouini, and G. E. Øien, “A threshold-based channel state feedback algorithm for modern cellular systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2422-2426, July 2007.

[6] X. Liu, E. K. P. Chong, and N. B. Shroff, “A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451-474, Mar. 2003.

[7] M. Andrews, L. Qian, and A. Stolyar, “Optimal utility based multi-user throughput allocation subject to throughput constraints," in *Proc. IEEE INFOCOM*, Miami, Mar. 2005, pp. 2415-2424.

[8] M.-S. Alouini and A. J. Goldsmith, “Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1165-1181, July 1999.

[9] Y. Kim and G. U. Hwang, “Threshold-based opportunistic scheduling for ergodic rate guarantees in wireless networks: Extended version." [Online] Available: http://queue.kaist.ac.kr/~guhwang, 2008.

[10] J. E. Marsden and A. J. Tromba, *Vector Calculus*, 4th ed. Freeman, 1996.

[11] Y. R. Zheng and C. Xiao, “Improved models for the generation of multiple uncorrelated Rayleigh fading waveforms," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 256-258, June 2002.