Opportunistic Downlink Transmission With Limited Feedback

Shahab Sanayei, *Member, IEEE*, and
Aria Nosratinia, *Senior Member, IEEE*

***Abstract*—Opportunistic scheduling provides attractive sum-rate capacities in a multiuser network when the base-station has transmit-side channel state information (CSI), which is often estimated at the mobiles and provided to the base station via a feedback channel. This correspondence investigates opportunistic methods in the presence of limited feedback. For flat Rayleigh-fading channels, strategies with only one-bit feedback per user are demonstrated that capture the double-logarithmic capacity growth (with number of users) of full-CSI systems. Furthermore, for a given system configuration, it is shown that if the one-bit feedback is chosen judiciously, there is little to be gained by increasing the feedback rate. Our results provide optimal methods of calculating the one-bit feedback, as well as expressions for the sum-rate capacity in the one-bit feedback regime. It is shown that one may achieve proportional fairness of scheduling in this regime with no loss of throughput. For OFDM multiuser systems, the motivation for limited feedback is even more pronounced. An extension of the one-bit technique is presented for subchannel/user selection under both correlated and uncorrelated subchannel conditions, and optimal growth in capacity is demonstrated.**

***Index Terms*—Capacity, channel state information (CSI), fairness, limited feedback, opportunistic communication.**

I. Introduction

In a multiuser environment it is highly probable that at least one link has high quality at any given point in time. Taking advantage of this opportunity leads to *multiuser diversity*. Obviously, multiuser diversity requires the base station to know the channel coefficients for all users, which are estimated at the mobiles and fed back to the base station. This information consists of real- or complex-valued variables that may require significant feedback rate. In the context of frequency-selective channels, multiple variables must be conveyed back to the transmitter, thus further increasing the feedback rate. This work demonstrates that it is possible to capture the multiuser diversity advantage even with a very low-rate feedback, in fact as low as one bit per data stream. We calculate the performance with limited feedback, and develop asymptotically optimal scheduling algorithms in the presence of limited feedback. We concentrate on single-beam opportunistic communication; extension to multiple-beams is relatively straight forward.

The notion of multiuser diversity is due to Knopp and Humblet [1] for the uplink, where they mentioned that the best strategy is to always transmit to the user with the best channel. Tse [2] provided a similar result for the downlink. Bender *et al.* [3] examined practical aspects of downlink multiuser diversity in the context of the IS-856 standard. Viswanath, Tse and Laroia [4] examined this problem for the downlink and presented a method of opportunistic beamforming via phase randomization. Hochwald, Marzetta, and Tarokh [5] investigate the problem of scheduling and rate feedback in the case of multiple-input–multiple-output (MIMO) channels. Sharif and Hassibi [6] generalized the opportunistic beamforming of [4] to the case where multiple beams are used simultaneously.

Some of works in the existing literature raise the question of the required feedback information, but to our knowledge, only [5]–[7] explicitly quantify the required feedback. In [7] the idea of thresholding was proposed for reducing the feedback load for exploiting multiuser diversity. Our method guarantees optimal growth rate with number of users via one-bit fixed-rate feedback, while [7] requires a variable-rate feedback of real-valued numbers and, to our understanding, it has not been proved to guarantee optimal growth rates. Similarly, the work in [5] does not consider capacity growth and does not minimize the feedback rate. The work in [6] reduces the number of real-valued variables that must be conveyed to the transmitter, but does not directly address the question of feedback rate (transmission of real-values requires infinite rate). In this correspondence we present a one-bit quantization strategy and the associated scheduling algorithm that guarantees optimal capacity growth rate. We demonstrate the viability of this approach in flat fading as well as frequency-selective channels. We also show that any additional feedback over and above one bit[1] does not deliver any significant improvement in sum-rate capacity. We investigate the fairness issues that are of concern in all opportunistic schemes. We present analytical tools that can be used for asymptotic capacity analysis of various opportunistic systems with limited feedback.

We use the following notation: $\mathbb{E}[\,]$ refers to the expected value of a random variable, $\gamma \approx 0.577$ is the Euler-Mascheroni constant. The asymptotic equivalence of two sequences a_n and b_n is denoted by $a_n \doteq b_n$, defined as $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. All capacities are in Nats/Sec/Hz.

II. System Model

We consider a multiuser cellular network with n users, all receiving data from the base station. We assume the block fading model for each user's channel. The channel state information of each user is assumed to be fully known to that user, and it is constant over a coherence interval of length T. The users and the base station are each equipped with one antenna. Under the block-fading frequency nonselective assumption, we have the following model for the received signal for each user:

$$y_i(t) = \sqrt{\rho_i} h_i s_i(t) + z_i(t). \tag{1}$$

In the above model, $s_i(t) \in \mathbb{C}^T$ is the vector of transmitted symbols of the ith user at time t with power constraint $\mathbb{E}\left[\|s_i(t)\|^2\right] = T$, and $y_i(t) \in \mathbb{C}^T$ is the received signal of the ith user at time t, $z_i(t) \sim \mathcal{CN}(0, I_T)$ is the independent and identically distributed (i.i.d.) complex Gaussian noise, h_i is the channel gain of the ith user, which is assumed to be zero mean circularly symmetric complex Gaussian random variable with unit variance per dimension. The users have mutually independent channel gains. Moreover we assume a homogeneous network in which all users have the same SNR, i.e., $\rho_i = \rho$. For each user there exits a low-rate but reliable and delay-free feedback channel to the base station.

III. Scheduling Via One-Bit Feedback

The base station sets a threshold α for all users. Each user compares the absolute value of their channel gain to this threshold. The *eligible users*, whose channel gains are above the threshold α, convey the 1-bit information about the quality of their channel to the base-station through a feedback channel. If the feedback channel is shared by

Manuscript received October 5, 2005; revised March 7, 2007.

S. Sanayei is with Huawei Technologies, Plano, TX 75075 USA (e-mail: sanayei@ieee.org).

A. Nosratinia is with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: aria@utdallas.edu).

Communicated by E. Modiano, Associate Editor for Communication Networks.

Color versions of Figures 1–5 in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2007.899529

[1]Subject to a judicious choice of threshold and smart scheduling, to be discussed in the sequel.

all users, a code division multiple access method can be used to convey the feedback bits to the base-station. When a dedicated low bit-rate feedback channels exits, each eligible user sends a "1" to indicate that its own channel gain is above the threshold. In this case, the base station recognizes the eligible users and then randomly picks an eligible user for transmission.[2] If the base station does not find any of the users eligible for transmission, then no signal is transmitted in that interval.[3]

A. The Sum-Rate Capacity

Upon receipt of each set of feedback bits, the base station transmits to users whose channel gain is above the threshold α (eligible users). Let $p = Pr\left[|h_i|^2 > \alpha\right] = e^{-\alpha}$. Since the channel gains are all mutually independent, the probability of having k eligible users is binomial, i.e.

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2}$$

The ergodic capacity upon receiving k ones by the base station is

$$\begin{aligned}\overline{C}_k &= \sum_{i=1}^{k} Pr[\text{the } i\text{th best user is selected}]\, C_i \\ &= \frac{1}{k}\sum_{i=1}^{k} C_i \end{aligned}\tag{3}$$

where $C_i = \int_0^\infty \log(1+\rho x) dF_i(x)$ and $F_i(x)$ is the CDF of the ith highest absolute value of all channel gains. In other words if $\{X_1, \dots, X_n\}$ is a permutation of $\left\{|h_1|^2, \dots, |h_n|^2\right\}$ such that $0 \le X_n \le \cdots \le X_1$, then $F_i(x) = Pr[X_i < x]$. When the channel gains are i.i.d., we have

$$F_i(x) = \sum_{l=0}^{i-1} \binom{n}{l} (F(x))^{n-l} (1-F(x))^l \tag{4}$$

where $F(x) = 1 - e^{-x}$ is the CDF of $|h_i|^2$ for $i = 1, \dots, n$. Thus, we can use the law of total probability to formulate the the sum-rate capacity of the network with one-bit feedback as follows:

$$C_{1\_bit} = \sum_{k=1}^{n} p_k \overline{C}_k = \mathbb{E}[\log(1+\rho X_\pi)] \tag{5}$$

where X_π is a random variable with the following probability cumulative distribution

$$F_\pi(x) = \sum_{k=1}^{n} p_k \overline{F}_k(x)$$

where

$$\overline{F}_k(x) = \frac{1}{k}\sum_{i=1}^{k} F_i(x).$$

B. The Optimal Threshold

The sum-rate capacity is a function of ρ, p and n. On the other hand the threshold α is uniquely determined by p from the following formula:

$$\alpha = F^{-1}(1-p). \tag{6}$$

For a Rayleigh-fading channel, the channel magnitude squared obeys an exponential law

$$\alpha = -\log p. \tag{7}$$

In order to find the optimal threshold we choose p such that the sum-rate capacity C_{1_bit} is maximized. The cost function $C_{1\_bit}(p)$ is a weighted sum of functions of the form $p^k(1-p)^{n-k}$ which are all concave over the interval $[0,1]$, hence C_{1_bit} is a concave function of p and it has a unique maximum over the interval $[0,1]$. To calculate the value of p that maximizes the sum-rate capacity, we must solve $\frac{\partial C_{1\_bit}(p)}{\partial p} = 0$ for p, i.e.,

$$\sum_{k=1}^{n} (k - np) p_k \overline{C}_k = 0. \tag{8}$$

A closed-form solution to this equation is in general not tractable. A numerical solution is possible with $O(n)$ complexity. We shall see that for asymptotic analysis, the exact value of this threshold is not needed.

C. Asymptotic Analysis of Sum-Rate Capacity

When channel state information is fully available at the base station, the base station only transmits to the user with the best channel, hence the ergodic sum-rate capacity of the network can be calculated by the following formula:

$$\begin{aligned} C_{full\_CSI} = C_1 &= \int_0^\infty \log(1+\rho x) dF_1 \\ &= n\int_0^\infty \log(1+\rho x) e^{-x}(1-e^{-x})^{n-1} dx \\ &\mathrel{\overset{\circ}{=}} \log(1+\rho\mu_1) \\ &\mathrel{\overset{\circ}{=}} \log(\log n) + \log \rho. \end{aligned}\tag{9}$$

where $\mathrel{\overset{\circ}{=}}$ indicates asymptotic equivalence, as defined earlier.

A natural question is: what is the loss in sum-rate due to a limited channel knowledge at the base station? The gap between the sum rate capacity of the fully informed network and the network with 1-bit CSI feedback, in the asymptote of large number of users, is illuminated via the following result.

Theorem 1: Consider a broadcast network (multiuser downlink) where the transmitter has one bit of CSI per user, indicating whether each user's channel magnitude is above or below a certain threshold α. With appropriate choice of α, the sum-rate loss incurred due to the 1-bit feedback vanishes in the asymptote of large number of users n. In other words

$$\Delta C = (C_{full\text{-}CSI} - C_{1-bit}) \underset{n\to\infty}{\longrightarrow} 0.$$

Proof: Equation (8) can be rewritten as

$$C_{1\_bit} = \frac{1}{np}\sum_{k=1}^{n} k p_k \overline{C}_k. \tag{10}$$

For a p satisfying (10) we have

$$\begin{aligned} C_{1\_bit} &= \frac{1}{np}\sum_{k=1}^{n} k p_k \overline{C}_k \\ &= \frac{1}{np}\sum_{k=1}^{n} k p_k \left(\frac{1}{k}\sum_{i=1}^{k} C_i\right) \\ &= \frac{1}{np}\sum_{k=1}^{n}\sum_{i=1}^{k} p_k C_i \\ &= \sum_{i=1}^{n}\left(\frac{1}{np}\sum_{k=i}^{n} p_k\right) C_i. \end{aligned}\tag{11}$$

[2]The scheduling to users with favorable channels may also be implemented via round robin. In the long run, both these strategies have the same average throughput per user. However, the round-robin version may be more appealing from a fairness point of view.

[3]In the absence of any users above threshold, the base station can also randomly pick a user for data transmission, although for large number of users this has vanishing advantage over no transmission.

Notice that $\pi_i = \frac{1}{np}\sum_{k=i}^{n} p_k,\ i = 1,\dots,n$ is a valid pmf because $\sum_{i=1}^{n}\pi_i = 1$, hence

$$\begin{aligned} C_{1\_bit} &= \sum_{i=1}^{n}\pi_i C_i \\ &= \sum_{i=1}^{n}\pi_i \int_0^\infty \log(1+\rho X)dF_i \\ &= \int_0^\infty \log(1+\rho x) d\left(\sum_{i=1}^{n}\pi_i F_i\right) \\ &= \int_0^\infty \log(1+\rho x)dF_\pi \end{aligned} \tag{12}$$

where F_π, under the optimality conditions mentioned above, can be alternatively written as

$$F_\pi = \sum_{i=1}^{n}\pi_i F_i \tag{13}$$

is a mixture probability measure of all order statistics of the exponential family. To proceed with the proof, we need to bound certain probabilities for this distribution, which in turn requires the following Lemma.

Lemma 1: For the distribution F_π mentioned above, we have

$$\frac{\sigma_\pi}{\mu_\pi}\xrightarrow[n\to\infty]{} 0$$

where μ_π and σ_π^2, are the mean and variance of F_π, respectively.

Proof:

$$\mu_\pi = \sum_{i=1}^{n}\pi_i\mu_i \tag{14}$$

where $\mu_i = \int_0^\infty x dF_i(x)$ is the mean of the ith order statistics of the exponential family, and

$$\begin{aligned} \sigma_\pi^2 &= \int_0^\infty (x-\mu_\pi)^2 dF_\pi(x) \\ &= \int_0^\infty x^2 dF_\pi(x) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i\int_0^\infty x^2 dF_i(x) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i\left(\sigma_i^2+\mu_i^2\right) - \mu_\pi^2 \\ &= \sum_{i=1}^{n}\pi_i\sigma_i^2 + \sum_{i=1}^{n}\pi_i\mu_i^2 - \mu_\pi^2 \end{aligned} \tag{15}$$

where $\sigma_i^2 = \int_0^\infty (x-\mu_i)^2 dF_i$ is the variance of the ith order statistics of the exponential family. It is a known fact (e.g., [8, Sec. 4.6]) that the mean and variance of the ordered exponential distributions $F(x) = 1-e^{-x}$ are given as follows:

$$\mu_i = \sum_{j=i}^{n}\frac{1}{j} = H_n - H_{i-1}$$

$$\sigma_i^2 = \sum_{j=i}^{n}\frac{1}{j^2} = S_n - S_{i-1}$$

where

$$H_k \triangleq \begin{cases}\sum_{j=1}^{k}\frac{1}{j}, & k>0\\ 0, & k=0\end{cases} \qquad S_k \triangleq \begin{cases}\sum_{j=1}^{k}\frac{1}{j^2}, & k>0\\ 0, & k=0.\end{cases}$$

It follows that

$$\begin{aligned} \mu_\pi &= \sum_{i=1}^{n}\pi_i\mu_i = \sum_{i=1}^{n}\pi_i(H_n - H_{i-1}) \\ &= H_n - \sum_{i=1}^{n}\pi_i H_{i-1} < H_n = \mu_1. \end{aligned} \tag{16}$$

It is known [9] that for all $k \ge 1$

$$\log k + \gamma + \frac{1}{2(k+1)} < H_k < \log k + \gamma + \frac{1}{2k} \tag{17}$$

using Jensen's inequality we have

$$\begin{aligned} \mu_\pi &= H_n - \sum_{i=1}^{n}\pi_i H_{i-1} \\ &> H_n - \sum_{i=1}^{n}\pi_i H_i \\ &> H_n - \gamma - \sum_{i=1}^{n}\pi_i\log i - \frac{1}{2}\sum_{i=1}^{n}\frac{\pi_i}{i} \\ &> H_n - \gamma - \log\left(\sum_{i=1}^{n} i\pi_i\right) - \frac{1}{2}\sum_{i=1}^{n}\pi_i \\ &> H_n - \gamma - \frac{1}{2} - \log\left(\sum_{i=1}^{n} i\pi_i\right) \end{aligned} \tag{18}$$

on the other hand

$$\begin{aligned} \sum_{i=1}^{n} i\pi_i &= \frac{1}{np}\sum_{i=1}^{n} i\sum_{k=i}^{n}p_k \\ &= \frac{1}{np}\sum_{k=1}^{n}p_k\sum_{i=1}^{k} i \\ &= \frac{1}{np}\sum_{k=1}^{n}p_k\left(\frac{k(k+1)}{2}\right) \\ &= \frac{\sum_{k=1}^{n}k^2p_k + \sum_{k=1}^{n}kp_k}{2np} \\ &= \frac{(n-1)p}{2} + 1 \end{aligned} \tag{19}$$

from (16), (18) and (19) we get

$$H_n - \log(np+2-p) - \gamma - \log(2\sqrt{e}) < \mu_\pi < H_n. \tag{20}$$

By inspecting (8) we also notice that $p_{opt} = O\left(\frac{1}{n}\right)$, because in order to have equality, the number of positive and negative terms in (8) should be of the same order in the asymptote of large n. Equivalently, the optimal threshold α scales logarithmically in the asymptote of large n (this fact can also be seen in Fig. 1 in which the X-axis is in logarithmic scale). Therefore, (20) suggests that

$$H_n - \mu_\pi = O(1) \tag{21}$$

or

$$\mu_\pi \doteq \mu_1 \doteq \log n \tag{22}$$

as $n \to \infty$. On the other hand

$$\begin{aligned} \sigma_\pi^2 &= \sum_{i=1}^{n}\pi_i\sigma_i^2 + \sum_{i=1}^{n}\pi_i\mu_i^2 - \mu_\pi^2 \\ &< \sigma_1^2 + \mu_1^2 - \mu_\pi^2 \\ &= \sigma_1^2 + (\mu_1+\mu_\pi)(\mu_1-\mu_\pi) \\ &< \sigma_1^2 + 2\mu_1(\mu_1-\mu_\pi) \end{aligned} \tag{23}$$

we also notice that $S_n < S_\infty = \frac{\pi^2}{6} < 2$ thus

$$\sigma_\pi^2 < 2 + 2\mu_1(\mu_1-\mu_\pi) \tag{24}$$

hence from (21), (22) and (24) we have

$$0 \le \left(\frac{\sigma_\pi}{\mu_\pi}\right)^2 < \frac{2}{\mu_\pi^2} + 2\left(\frac{\mu_1}{\mu_\pi}\right)\left(\frac{\mu_1}{\mu_\pi} - 1\right) \to 0 \tag{25}$$

as $n \to \infty$. Thus (22) implies $\frac{\sigma_\pi}{\mu_\pi} \to 0$ as $n \to \infty$. □

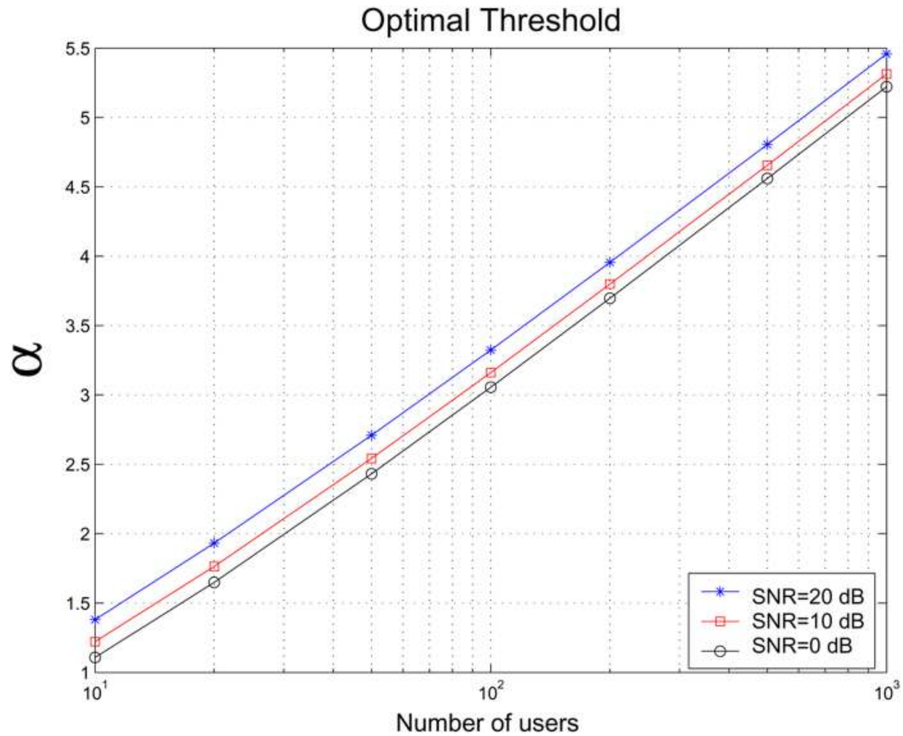

Fig. 1. Optimal threshold versus number of users for different SNR values.

Let X_π be the random variable associated with the probability measure F_π, then using the Chebyshev's inequality for all $\epsilon > 0$ we have

$$\begin{aligned}\Pr\left[\left|\frac{1+\rho X_\pi}{1+\rho\mu_\pi} - 1\right| > \epsilon\right] &= \Pr\left[\left|\frac{X_\pi - \mu_\pi}{1/\rho + \mu_\pi}\right| > \epsilon\right] \\ &\le \frac{\mathbb{E}[(X_\pi - \mu_\pi)^2]}{\epsilon^2(1/\rho + \mu_\pi)^2} \\ &\le \frac{1}{\epsilon^2}\left(\frac{\sigma_\pi}{\mu_\pi}\right)^2\end{aligned}$$

hence $\frac{1+\rho X_\pi}{1+\rho\mu_\pi} \xrightarrow{i.p.} 1$. Using the continuous mapping theorem [10], we have

$$\log\left(\frac{1+\rho X_\pi}{1+\rho\mu_\pi}\right) \xrightarrow{i.p.} 0. \tag{26}$$

Now to conclude the proof it is enough to show that the random variable

$$Y_n = \log\left(\frac{1+\rho X_\pi}{1+\rho\mu_\pi}\right) \tag{27}$$

is uniformly integrable. According to Lemma 2 (Appendix I) it is enough to show that

$$\lim_{c\to\infty}\limsup_n \int_c^\infty \Pr[|Y_n| > y]dy = 0.$$

We note that because X_n has exponential tail and well behaved at 0, for all n, $\mathbb{E}[|Y_n|] < \infty$. We have

$$\int_c^\infty \Pr[|Y_n| > y]dy = I_+ + I_-$$

where

$$\begin{aligned}I_+ &= \int_c^\infty \Pr[Y_n > y]dy, \\ I_- &= \int_c^\infty \Pr[Y_n < -y]dy.\end{aligned}$$

Using the Markov inequality we have

$$\begin{aligned}I_+ &= \int_c^\infty \Pr[(1+\rho X_\pi) > e^y(1+\rho\mu_\pi)]dy \\ &\le \frac{\mathbb{E}[1+\rho X_\pi]}{1+\rho\mu_\pi}\int_c^\infty e^{-y}dy = e^{-c}.\end{aligned} \tag{28}$$

On the other hand

$$\begin{aligned}I_- &= \int_c^\infty \Pr[(1+\rho X_\pi) < e^{-y}(1+\rho\mu_\pi)]dy \\ &= \int_c^\infty \Pr\left[X_\pi < \frac{e^{-y}(1+\rho\mu_\pi)-1}{\rho}\right]dy\end{aligned}$$

since almost surely $X_\pi \ge 0$, we have $y \le \log(1+\rho\mu_\pi)$ thus the upper limit of integral can be replaced by $\log(1+\rho\mu_\pi)$

$$I_- = \int_c^{\log(1+\rho\mu_\pi)} \Pr\left[X_\pi < \frac{e^{-y}(1+\rho\mu_\pi)-1}{\rho}\right]dy$$

After change of variable $u = \frac{e^{-y}(1+\rho\mu_\pi)-1}{\rho}$ we have

$$I_- = \int_0^{\beta_n} \frac{\Pr[X_\pi < u]}{1/\rho + u}du$$

where $\beta_n = \frac{e^{-c}(1+\rho\mu_\pi)-1}{\rho}$. Using (17) and (20), we have $\beta_n < \overline{\beta}_n$ where $\overline{\beta}_n = a\log n + b$, $a = e^{-c}$, and $b = \frac{1-e^{-c}}{\rho} + e^{-c}(1+\gamma)$. Hence

$$\begin{aligned}I_- &\le \int_0^{\overline{\beta}_n} \frac{\Pr[X_\pi < u]}{1/\rho + u}du \\ &\le \Pr[X_\pi < \overline{\beta}_n]\int_0^{\overline{\beta}_n}\frac{du}{1/\rho + u} \\ &= \log(1+\rho\overline{\beta}_n)\cdot\Pr[X_\pi < \overline{\beta}_n].\end{aligned}$$

It is sufficient to show that $\Pr[X_\pi < \overline{\beta}_n] \to 0$ faster than $\frac{1}{\log\log n}$. From (4) we have

$$\begin{aligned} F_i(x) &= \sum_{l=0}^{i-1} \binom{n}{l} (1-e^{-x})^{n-l} e^{-lx} \\ &\le (1-e^{-x})^{n-i+1} \sum_{l=0}^{i-1} (ne^{-x})^l \\ &= (1-e^{-x})^{n-i+1} \frac{(ne^{-x})^i - 1}{ne^{-x}-1} \\ &< (1-e^{-x})^{n-i+1} \frac{(ne^{-x})^i}{ne^{-x}-1} \\ &= \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \left(\frac{ne^{-x}}{1-e^{-x}}\right)^i \end{aligned}$$

where we have used $\binom{n}{l} \le n^l$. It follows

$$\begin{aligned} &\overline{F}_k(x) = \frac{1}{k}\sum_{i=1}^{k} F_i(x) \le \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \frac{1}{k} \sum_{i=1}^{k} \left(\frac{ne^{-x}}{1-e^{-x}}\right)^i \\ &= \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \frac{1}{k} \left(\frac{ne^{-x}}{(n+1)e^{-x}-1}\right) \left[\left(\frac{ne^{-x}}{1-e^{-x}}\right)^k - 1\right] \\ &< \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \cdot \left(\frac{ne^{-x}}{1-e^{-x}}\right)^k. \end{aligned}$$

Now we can bound the CDF of X_π

$$\begin{aligned} F_\pi(x) &= \sum_{k=1}^{n} p_k \overline{F}_k(x) \\ &< \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \cdot \sum_{k=1}^{n} p_k \left(\frac{ne^{-x}}{1-e^{-x}}\right)^k \\ &< \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \cdot \sum_{k=0}^{n} p_k \left(\frac{ne^{-x}}{1-e^{-x}}\right)^k. \end{aligned}$$

We know the *moment generating function* of the binomial distribution is

$$\mathbb{E}[z^k] = \sum_{k=0}^{n} p_k z^k = (1-p+pz)^n.$$

Using the fact, we substitute $z = \frac{ne^{-x}}{1-e^{-x}}$ to arrive at

$$\begin{aligned} F_\pi(x) &< \frac{(1-e^{-x})^{n+1}}{ne^{-x}-1} \left(1 - p + \frac{np\, e^{-x}}{1-e^{-x}}\right)^n \\ &< \frac{1}{ne^{-x}-1} \left(1 - p + \frac{np}{e^x - 1}\right)^n. \end{aligned}$$

Let $x = \overline{\beta}_n = a\log n + b$ and $p = \lambda/n$ where $\lambda = \sum_{k=1}^{n} kp_k$ is the *average number of eligible users*, we have[4]

$$\begin{aligned} I_- &< \log(1+\rho\overline{\beta}_n) \cdot F_\pi(\overline{\beta}_n) \\ &< \frac{\log(\rho a \log n + \rho b + 1)}{e^b n^{1-a} - 1} \left(1 - \frac{\lambda}{n} + \frac{\lambda}{e^b n^a - 1}\right)^n. \end{aligned}$$

Because $a = e^{-c} < 1$

$$\frac{\log(\rho a\log n + \rho b + 1)}{e^b n^{1-a} - 1} \to 0$$

[4]We use the fact that $F(\cdot)$ is an exponential CDF. For the general non-Rayleigh fading case, as long as $F(\cdot)$ has exponentially decaying tail (e.g., Rician or m-Nakagami), one can use the *extreme value theory* and a similar argument to generalize our result to non-Rayleigh fading case. The details are beyond the scsope of this correspondence.

and

$$\left(1 - \frac{\lambda}{n} + \frac{\lambda}{e^b n^a - 1}\right)^n \to e^{-\lambda}$$

as $n \to \infty$ therefore, we can conclude $I_- \to 0$ as $n \to \infty$.

Combining this with the counterpart inequality (28) we have

$$\limsup_n \int_c^\infty \Pr[|Y_n| > y]dy = \limsup_n (I_+ + I_-) \le e^{-c}$$

and finally we conclude

$$\lim_{c\to\infty} \limsup_n \int_c^\infty \Pr[|Y_n| > y]dy = 0.$$

Hence, Y_n is uniformly integrable and therefore by Theorem 2, we conclude that $\lim_{n\to\infty} \mathbb{E}[Y_n] = 0$. This implies

$$C_{1\_bit} - \log(1+\rho\mu_\pi) \to 0,$$

similarly, we can show that

$$C_{full\text{-}CSI} - \log(1+\rho\mu_1) \to 0$$

as $n \to \infty$. Because of (22), we have $\log\left(\frac{1+\rho\mu_1}{1+\rho\mu_\pi}\right) \to 0$, hence we can conclude

$$\Delta C = C_{full\text{-}CSI} - C_{1\_bit} \to 0$$

as $n \to \infty$. □

Corollary 1: If $p = \frac{\lambda}{n^\kappa}$ for any fixed $\kappa \ge 1$, the sum-rate difference between 1-bit feedback and full CSI asymptotically vanishes.

This means that the capacity scaling laws of 1-bit feedback is as good as the full-CSI system (with vanishing difference) as long as we choose the threshold α to be an affine function of $\log n$. In particular, we can optimize the parameters of this affine function to maximize the rate (c.f. (8) and Fig. 1). Also note that λ (the average number of eligible users) can be used as a design parameter for tradeoff between rate and fairness.

D. Simulation Results

Fig. 2 shows the sum-rate capacity of a single-input–single-output (SISO) network. As it can be seen in the figure, our proposed scheduling, with only 1-bit feedback, has the same double logarithmic growth rate as the fully informed network. The capacity loss is minimal and is expected to vanish for very large n (number of users). For a practical range of n, scheduling with 1-bit feedback also captures most of the capacity of the fully informed network for a wide range of SNR values. The gap between the two curves closes at very high values of n, due to sublogarithmic convergence. Fig. 1 shows the optimal threshold for various of SNR values. It can be seen that the optimal threshold scales logarithmically with number of users (in Fig. 1 the x-axis is in logarithmic scale).

IV. Fairness

Opportunistic scheduling increases the overall throughput of the system, but then the regularity of round-robin scheduling is relinquished. In general, there is no bound on the delay of a user while it waits to be serviced, which has practical drawbacks. To address this concern, the concept of *proportional fairness* has been introduced [4], [11].

Proportionally fair (PF) scheduling provides to each user a share of transmission time proportional to the achievable throughput of that user. This achievable throughput is measured causally over a fixed

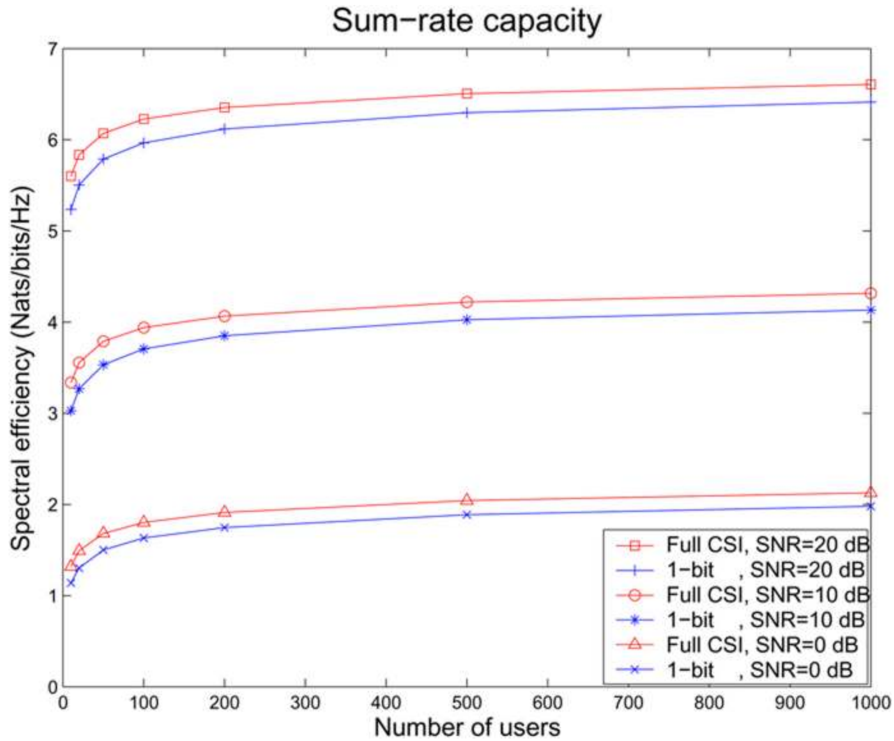

Fig. 2. Comparison of sum-rate capacity for 1-bit and full CSI scheduling for different values of SNR.

window of observation. Assume that in time slot t, user k's channel can support rate $R_k(t)$. Assuming that the feedback is instantaneous and error-free, the scheduler assigns the time slot t to the user k^* such that

$$k^* = \arg\max_k \frac{R_k(t)}{T_k(t)}$$

the rate $T_k(t)$ is the kth user's average throughput in a window of size t_c and it is calculated as follows [4]

$$T_k(t+1) = \begin{cases} \left(1 - \frac{1}{t_c}\right) T_k(t) + \frac{1}{t_c} R_k(t), & k = k^* \\ \left(1 - \frac{1}{t_c}\right) T_k(t), & k \neq k^*. \end{cases} \tag{29}$$

The window t_c is usually chosen to be much greater than the average small-scale fading coherence time of users, in IS-856, $t_c = 1.67$ second.

The 1-bit scheduling method as proposed in Section III, with random selection among eligible users, does not provide proportional fairness. However, if we perform the proportionally fair algorithm mentioned above by scheduling among *eligible users*, then proportional fairness is achieved. In other words, we choose the user k^* such that

$$k^* = \arg\max_{k \in \mathcal{E}} \frac{R_k(t)}{T_k(t)} \tag{30}$$

where

$$\mathcal{E} = \left\{ k : |h_k|^2 > \alpha \right\}$$

Fig. 3 shows the empirical CDF of the rate assigned to a user at SNR $=$ 10 dB. As can be seen the distribution of the random selection scheduling and Max-SNR (always allocating the channel to the users with the highest channel gain) have much higher spread compared to the proportionally fair scheduling for both 1-bit and full-CSI cases. The plots also indicates that, as the number of users increases, the difference between the PF scheduling with one-bit and full-CSI becomes negligible. The capacity plots depicted in Fig. 4 show that we do not lose any throughput by achieving proportional fairness.

V. Opportunistic Multiuser OFDM With Limited Feedback

One of the major challenges in employing an opportunistic scheme in OFDM networks is the large amount of feedback required to the base-station. For example in 802.11a each user has 64 subchannels and a network of 100 users requires the base station to collect 6400 real numbers from all the users. To address this issue, Svedman *et al.* [12] proposed an opportunistic scheme in which adjacent subchannels are clustered into groups and then only the average SNR value of each cluster is fed back to the base station. But this still requires feeding back several real numbers to the base station. In this section we consider an extension of our results to OFDM multiuser networks, resulting in very good asymptotic performance with much smaller feedback rates.

For each user in the network, consider a frequency selective linear time invariant model

$$y_{t,k} = \sum_{i=0}^{\nu} h_{i,k} x_{t-i,k} + w_{t,k} \tag{31}$$

where $x_{t,k}$ and $y_{t,k}$ are the input and the output for the kth user ($k \in \{1, \ldots, K\}$) at time t, respectively, w is the additive white complex Gaussian noise and uncorrelated among the users with zero mean and variance σ_w^2, $h_{i,k}$ is the ith channel tap for user k and is distributed as $\mathcal{CN}(0,1)$ which is assumed to be uncorrelated among different users, although for each user, channel taps may or may not be correlated. ν is the memory of the channel and it is assumed to be the same for all users. We assume that the base-station uses OFDM for data transmission to each user. By applying cyclic prefix and IDFT, user k's channel is divided into N different subchannels $H_{n,k}$ such that

$$H_{n,k} = \frac{1}{\sqrt{N}} \sum_{t=0}^{\nu} h_{t,k} e^{-j\frac{2\pi n t}{N}}. \tag{32}$$

We also assume that the total transmission power in the network is limited by $P_{\max}$.

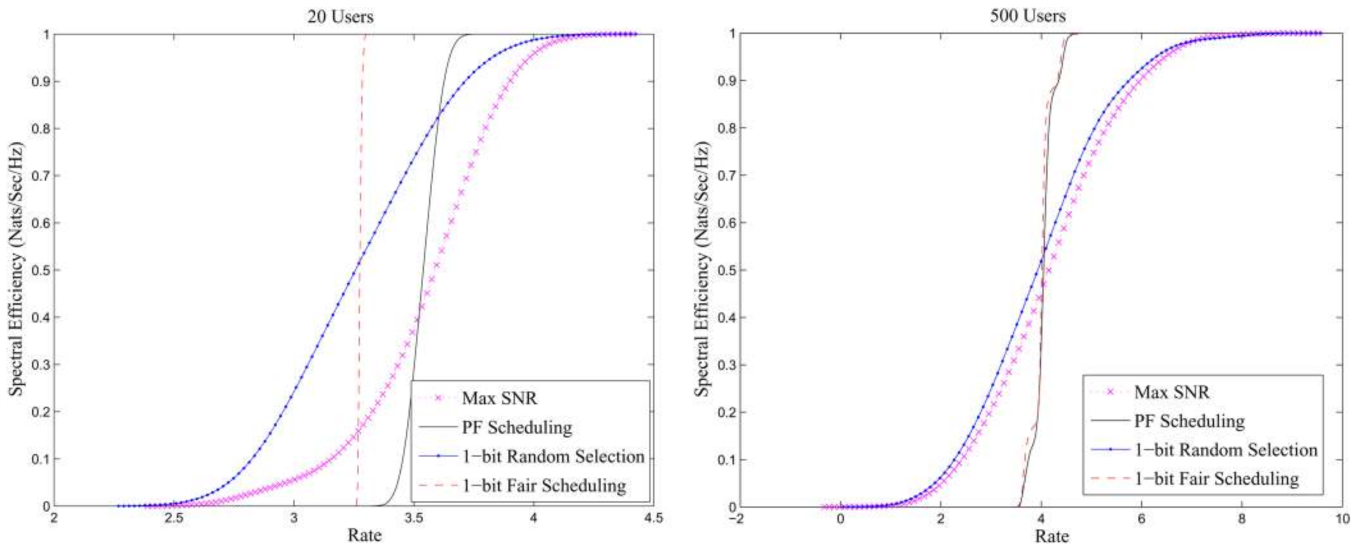

Fig. 3. Empirical distribution of rate at SNR = 10 dB for $n = 20$ and $n = 500$.

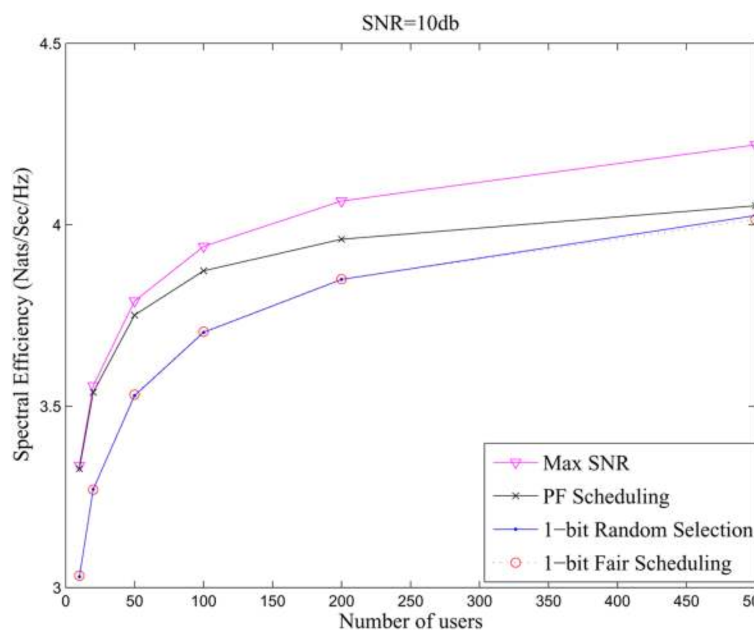

Fig. 4. Capacity at SNR = 10 dB.

When all users share the same bandwidth, and the base station has full information about every user's subchannels, then in order to maximize the sum-rate capacity of the network, the problem of subcarrier and power allocation to different users in the network must be solved jointly. However, this imposes a huge computational complexity at the base station. Especially if the wireless channel varies quickly, the base-station requires an enormous computational power to rapidly compute the optimal solution for power and subchannel allocation among the users. Moreover the optimal dynamic joint power and subchannel allocation requires fast and reliable feed-forward and feedback channels for exchanging information between the users and the base station. Especially with large number of users in the network, sending this information back and forth between the users and the base station causes a huge overhead for the network. This motivates a low-complexity sub-optimum algorithm.

One may achieve economy of computation and communication through separation of subchannel and power allocation. It is possible to first select subchannels and then perform water-filling among all selected subchannels, but this again requires the base station to send back the optimal power allocation vector to all the users, together with the indices of their selected subchannels. Yet another suboptimal scheme is to equally allocated the total power among all subchannels and then perform the subchannel allocation among all users [13]. We adopt the latter approach in this correspondence.

Assuming full channel knowledge at the base station, maximizing the sum-rate capacity of the network reduces to allocating subchannel to users that have the best channel conditions. In order to avoid intercarrier interference we allocate each frequency bin to a single user. Under this condition, maximum sum-rate capacity with equal power splitting among the subchannels is achieved when for each frequency bin we choose the user whose corresponding subchannel gain is maximum within that frequency band. The sum rate capacity in this case is given by

$$C_{full\_CSI} = \sum_{n=1}^{N} \log\left(1 + \mathrm{SNR}\cdot \max_k |H_{n,k}|^2\right) \tag{33}$$

where $\text{SNR} = \frac{P_{\max}}{N\sigma_w^2}$ is the SNR per subchannel. This subchannel selection scheme is in fact a generalization of the opportunistic scheduling in flat-fading multiuser networks [1] over N different flat fading subchannels provided by OFDM.

A. Subchannel Allocation With Limited Feedback

The opportunistic scheme mentioned in the previous section and most of the similar subchannel allocation schemes [13] require full knowledge of the subchannel information to be available at the transmitter. However, from a practical point of view this is not affordable because it requires a sum total of KN positive real numbers to be reliably transmitted to the base station, which is not affordable in practice. Svedman et. al [12] propose to divide each user's subchannels into clusters. Then in each cluster, the maximum value of the cluster is fed back to the base station. This reduces the number of real values to KL assuming that there are L clusters. But this still requires feeding back several real numbers to the base station without any error and delay which is still not attractive from an implementation point of view.

We propose a simple scheme where, instead of feeding back the full information of the subchannels, only one-bit of information per subchannel (or cluster) is fed back to the base station for subchannel allocation. For user k, the nth subchannel gain $|H_{n,k}|^2$ is compared to a threshold α_n, if the subchannel gain is above the threshold a 1 is transmitted back to the base station otherwise a 0 is transmitted.[5] So at most N bits per user is required in feedback.[6] Upon receipt of all feedback bits from the users, the base station allocates each subchannel to one of the users whose corresponding feedback bit is 1. This assignment can be done via either random selection or round robin scheduling among eligible users. Our claim is that by judicious choice of the threshold levels $\{\alpha_n\}$, most of the multiuser capacity gain is preserved. The choice of α_n's for each subchannel can be done according to Section III-B.

When channel taps are uncorrelated, i.e., $\mathbb{E}\left[h_{t,k}h_{s,k}^*\right] = \delta_{t-s}$, we can use the analytical framework developed in Section III for determining the optimal threshold value and the evaluation of the sum-rate capacity. We notice that under the assumption of uncorrelated channel taps, the subchannel gains $\{|H_{n,k}|^2\}$ are identically distributed exponential random variables. Hence the ergodic sum-rate capacity with full CSI at the base station is

$$C_{\text{full-CSI}} = N\mathbb{E}\left[\log\left(1 + \text{SNR} \max_k |H_{n,k}|^2\right)\right] \tag{34}$$

therefore we can apply Theorem 1 in each subchannel to show that

$$\Delta C = C_{\text{full-CSI}} - C_{\text{1-bit}} \longrightarrow 0$$

as $n \to \infty$.

B. Subchannel Correlation

Now we assume that for each user the channel taps are correlated, but there is no dependence between different users' channels. The correlation model that we consider is an exponential decaying model described by

$$\mathbb{E}\left[h_{t,k}h_{s,k}^*\right] = \rho^{|t-s|}. \tag{35}$$

[5]For contention-based feedback channels, a slight variation can lead to better channel utilization: When a subchannel is below the threshold level, instead of sending a 0, no feedback is transmitted.

[6]One can also use the idea of clustering the subchannels to reduce the amount of feedback to L bits per user.

Let $\eta_{n,k}$ be the power of the nth subchannel of the kth user which can be calculated as

$$\begin{aligned}
\eta_{n,k} &= \mathbb{E}\left[|H_{n,k}|^2\right] \\
&= \frac{1}{N}\mathbb{E}\left[\left(\sum_{t=0}^{N-1} h_{t,k}e^{-j\frac{2\pi tn}{N}}\right)\left(\sum_{s=0}^{N-1} h_{s,k}e^{-j\frac{2\pi sn}{N}}\right)^*\right] \\
&= \frac{1}{N}\mathbb{E}\left[\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} h_{t,k}h_{s,k}^* e^{-j\frac{2\pi(t-s)n}{N}}\right] \\
&= \frac{1}{N}\sum_{t=0}^{N-1}\sum_{s=0}^{N-1} \rho^{|t-s|}e^{-j\frac{2\pi n}{N}(t-s)}
\end{aligned}$$

by change of summing index to $u = t - s$ we get

$$\begin{aligned}
\eta_{n,k} &= \sum_{u=-(N-1)}^{N-1}\left(1 - \frac{|u|}{N}\right)\rho^{|u|}e^{-j\frac{2\pi n}{N}u} \\
&= \Re\{\beta_n\} - 1
\end{aligned} \tag{36}$$

where $\beta_n = \frac{1}{\sqrt{N}}\sum_{m=1}^{N-1} b_m e^{-j\frac{2\pi n}{N}m}$ is the discrete Fourier transform (DFT) of the sequence $b_m = \left(\sqrt{N} - \frac{m}{\sqrt{N}}\right)\rho^m$. Assuming $\rho^N \ll 1$, after some algebra we obtain the following expression for $\eta_{n,k}$:

$$\eta_{n,k} \approx \frac{1-\rho^2}{1 + 2\rho\cos\theta_n + \rho^2} + \frac{2\rho\cos\theta_n + 4\rho^2 + 2\rho^3\cos\theta_n}{N(1 + 2\rho\cos\theta_n + \rho^2)^2} \tag{37}$$

where $\theta_n = \frac{2\pi n}{N}$. Notice that for a given n, $\{H_{n,k}\}$'s are i.i.d. across different users, hence $\eta_{n,k}$ does not depend on k. Thus correlation between taps leads to subchannels with different qualities. On the other hand exact calculation of the optimal threshold for this case is not mathematically tractable. So we propose a suboptimal solution for quantizing the subchannels with one bit. For the nth frequency bin, we divide the subchannel gains by $\eta_{n,k}$ and then compare the normalized channel gain by the optimal threshold calculated in Section III-B, if

$$\frac{|H_{n,k}|^2}{\eta_{n,k}} \geq \alpha_n$$

the feedback bit is set to 1, otherwise it is set to 0.

Fig. 5 is the simulation result based on the proposed algorithm for $\text{SNR} = 10$ dB. As can be seen in the figure for both uncorrelated and correlated cases, the sum-rate capacity of our scheme closely follows the sum-rate capacity of the opportunistic subchannel selection. Moreover the capacity achieved by our scheme is much higher than TDMA scheduling and only slightly lower than the full CSI sum-rate capacity.

VI. Conclusion

In this correspondence, we investigate the performance of opportunistic multiuser systems in the limited-feedback regime. The developments in this correspondence show that with only one-bit feedback, most of the sum-rate capacity of the fully informed system can be achieved. Thus if the feedback bit is chosen judiciously, there is little to be gained by allocating any further feedback rate. We calculate the optimal thresholds to generate the one-bit feedback, and calculate the sum-rate capacity in this regime. It is possible to maintain proportional fairness without any loss of throughput in this regime. We then extend the results to frequency-selective channels via a simple joint user/subchannel selection strategy. Future work may address extension of these methods to multiple antenna systems as well as investigation of the robustness of these methods in the presence of channel estimation errors and feedback delay.

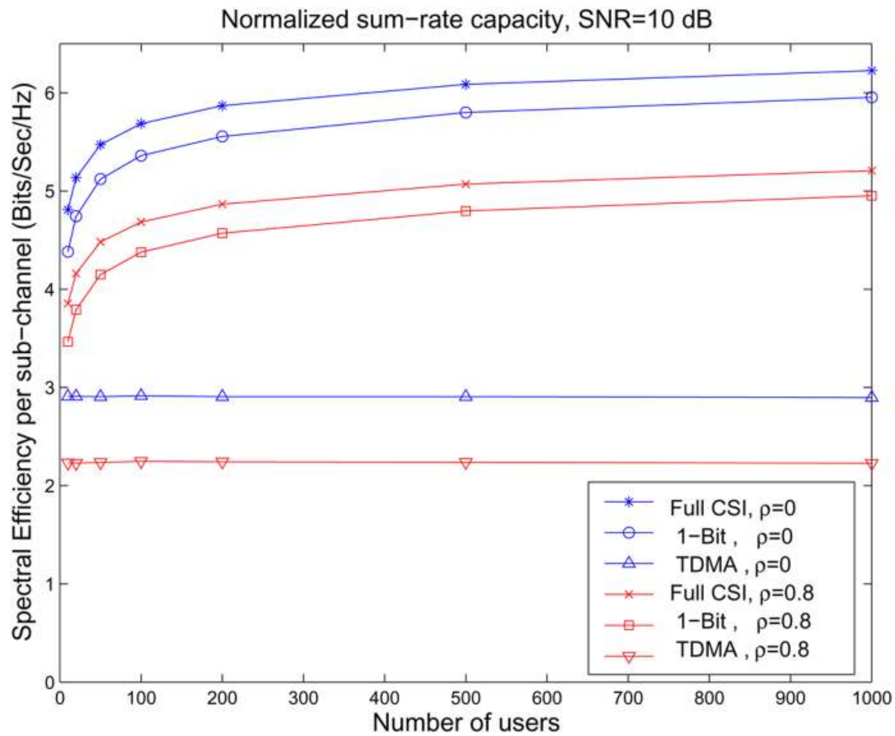

Fig. 5. Sum-rate capacity (normalized by N) versus number of users.

APPENDIX

Definition 1: The family of random variables $\{X_n\}$ is said to be uniformly integrable if

$$\lim_{c\to\infty}\limsup_{n}\int_{c}^{\infty} x\, dF_{|X_n|}(x) = 0.$$

Uniform integrability is a sufficient condition for convergence in mean.

Theorem 2: If X_n is a uniformly integrable random variable and $\mathbb{E}[X] < \infty$, then convergence in distribution, implies convergence in mean, i.e., if $X_n \xrightarrow{i.p.} X$ then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Proof: See [10]. □

The following lemma provides an alternative way to prove the uniform integrability.

Lemma 2: If $\mathbb{E}[|X_n|] < \infty$ for all n, then X_n is uniformly integrable if

$$\lim_{c\to\infty}\limsup_{n}\int_{c}^{\infty} \Pr[|X_n| > t]dt = 0.$$

Proof: For every n and $c > 0$ we have

$$c(1 - F_{|X_n|}(c)) \le \int_{c}^{\infty} x\, dF_{|X_n|}(x) \le \mathbb{E}[|X_n|] < \infty$$

hence $\lim_{c\to\infty} c(1 - F_{|X_n|}(c)) = 0$. Using integration by parts, we have

$$\begin{aligned} 0 &\le \int_{c}^{\infty} x\, dF_{|X_n|}(x) \\ &= -x(1 - F_{|X_n|}(x))\Big|_{c}^{\infty} + \int_{c}^{\infty} \Pr[|X_n| > x]\, dx \\ &= -c(1 - F_{|X_n|}(c)) + \int_{c}^{\infty} \Pr[|X_n| > x]\, dx \\ &\le \int_{c}^{\infty} \Pr[|X_n| > x]\, dx \end{aligned}$$

and this proves the lemma. □

Lemma 3: If $X_n \xrightarrow{i.p.} 0$ and $a_n \to 0$, then, $a_n X_n \xrightarrow{i.p.} 0$

Proof: For every $\epsilon, \delta_1, \delta_2 > 0$, there exits N_1 such that for all for all $n > N_1$ we have $|a_n| < \delta_1$. Also there exits N_2 such that for all $n > N_2$, $\Pr\left[|X_n| > \frac{\epsilon}{\delta_1}\right] < \delta_2$, thus for all $n > N = \min\{N_1, N_2\}$

$$\begin{aligned} Pr[|a_n X_n| > \epsilon] &= \Pr\left[|X_n| > \frac{\epsilon}{|a_n|}\right] \\ &\le \Pr\left[|X_n| > \frac{\epsilon}{\delta_1}\right] \le \delta_2 \end{aligned}$$

thus $a_n X_n \xrightarrow{i.p.} 0$. □

REFERENCES

[1] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. Int. Conf. Commun.*, Seattle, WA, Jun. 1995.

[2] D. Tse, "Optimal power allocation over parallel Gaussian channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Ulm, Germany, Jun. 1997.

[3] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR a bandwidth efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, pp. 70–77, Jul. 2000.

[4] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[5] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[6] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[7] D. Gesbert and S. Alouini, "How much feedback is multiuser diversity really worth," in *Proc. Int. Conf. Commun.*, Paris, France, Jun. 2004.

[8] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. New York: Wiley, 1992.
[9] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge, U.K.: Cambridge University Press, 1996.
[10] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.
[11] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR: A high efficiency-high data rate personal communication wireless system," in *Proc. Veh. Technol. Conf.*, Tokyo, Japan, May 2000, vol. 3, pp. 1854–1858.
[12] P. Svedman, S. K. Wilson, L. Cimini, and B. Ottersten, "A simplified opportunistic feedback and scheduling scheme for OFDM," in *Proc. Veh. Technol. Conf.*, Milan, Italy, May 2004.
[13] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. Veh. Technol. Conf.*, May 2000, vol. 2, pp. 1085–1089.

Fountain Capacity

Shlomo Shamai (Shitz), *Fellow, IEEE*,
İ. Emre Telatar, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—Fountain codes are currently employed for reliable and efficient transmission of information via erasure channels with unknown erasure rates. This correspondence introduces the notion of fountain capacity for arbitrary channels. In contrast to the conventional definition of rate, in the fountain setup the definition of rate penalizes the reception of symbols by the receiver rather than their transmission. Fountain capacity measures the maximum rate compatible with reliable reception regardless of the erasure pattern. We show that fountain capacity and Shannon capacity are equal for stationary memoryless channels. In contrast, Shannon capacity may exceed fountain capacity if the channel has memory or is not stationary.**

***Index Terms*—Arbitrarily varying channels, channel capacity, content distribution, erasure channels, fountain codes.**

I. Introduction

Fountain codes are a class of sparse-graph codes that have received considerable attention in the last few years. The first *fountain codes* were the LT erasure-correcting codes introduced by Luby in [1]. The LT codes are linear rateless codes that encode a vector of k symbols of information with an infinite sequence of parity-check bits. The parity-check equations (known to the decoder) are chosen equiprobably from a random ensemble: The cardinality of the parity-check equations has a histogram given by the so-called robust soliton distribution and all k information symbols have identical probability to participate in any given parity-check equation. The infinite sequence is transmitted through an erasure channel. The decoder runs a belief propagation algorithm observing only as many channel outputs as necessary to recover the k transmitted bits.

Manuscript received October 8, 2006; revised July 20, 2007. This work was supported in part by the National Science Foundation under Grants CCR-0312879 and CCF-0635154 and by the U.S.–Israel Binational Science Foundation under Grant 2004140. The material in this correspondence was presented at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006.

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa, 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

İ. E. Telatar is with the Information Theory Laboratory (LTHI), School of Computer and Communication Sciences (I &C), Ecole Polytechnique Federale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland (e-mail: Emre.Telatar@epfl.ch).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2007.907495

Better performance can be obtained with the fountain codes known as *raptor codes* introduced by Shokrollahi in [2] for erasure correction. Raptor codes have been applied to other channels such as binary channels in [3]–[5] and Gaussian channels [6].

A typical application of fountain codes is a system where the same message is to be broadcast simultaneously to several receivers, served by erasure channels with different erasure rates. The conventional Shannon-theoretic approach to this scenario is the compound channel (see, e.g., [7]), where the actual channel is unknown to the encoder and chosen from a given uncertainty set. Ensuring reliable communication for all receivers, the compound capacity is upper-bounded by the smallest capacity among those channels in the uncertainty set. This bound is tight in those cases, such as the compound erasure channel, in which the mutual information of all channels in the uncertainty class is maximized by the same input distribution. This setup not only requires the transmitter to cater to the worst channel conditions but it incurs a considerable waste of channel resources for those receivers that enjoy better erasure rates than the worst. The use of fountain codes enables receivers to stop listening to the channel once the information is decoded reliably. Thus, receivers only need to obtain from the channel a number of symbols that is a small multiple (close to 1) of the number of information symbols. This happens sooner for those receivers that experience favorable channel conditions. As customary in the information theory of channels with nonprobabilistic description of the uncertainty, we adopt a worst case approach in order to capture the robustness of the fountain codes with respect to the patterns of erasures.

Fountain codes have been adopted in the 3GPP wireless standard for Multimedia Broadcast/Multicast [8], [9] and they have been used in lossless data compression in [10].

In addition to their appealing conceptual structure, the commercial success and excellent efficiency achieved by fountain codes are incentives to investigate their Shannon-theoretic limits. The main difference from the standard Shannon setup is in the definition of rate: a fountain code is rateless (or zero-rate) in that it adds an infinite amount of redundancy to the information vector. Instead of defining the rate from the perspective of the encoder, in the fountain setup we define it from the perspective of the decoder: ratio of information symbols transmitted to channel symbols received. So while the classical definition of rate penalizes the use of the channel by the transmitter ("pay-per-use"), in the fountain setup the definition of rate penalizes the reception of (non-erased) symbols by the receiver ("pay-per-view"). Independent of the fountain code setting and within the context of broadcast channels, it has been recognized in [11]–[13] that the classical definition of rate is overly pessimistic for asynchronous broadcast where a common message is transmitted to several receivers which are "turned on" at not necessarily identical times. In [11]–[13], the individual rates in the capacity region are normalized by the time until the corresponding receiver is switched off. Recent works that deal with the conventional Shannon capacity of the concatenation of noisy channels and erasure channels include [14], [15].

This correspondence is organized as follows. In Section II, we give the definition of fountain capacity for an arbitrary channel, along with the associated notions of reliability and allowable encoding strategies. We show that fountain capacity is upper-bounded by Shannon capacity.