# Opportunistic Beamforming with Limited Feedback

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

***Abstract*— This work investigates the following question: subject to strictly limited (finite-rate) feedback in a multi-user multi-antenna system, what channel state information (CSI) should we send back to the transmitter, and how should it be used? Considering the class of single-beam systems, we suggest a combination of beamforming (array gain) and multi-user diversity. It has been shown that in single antenna systems, one bit of feedback per user can capture almost all gains available due to multi-user diversity, therefore we propose and analyze a compound strategy that uses one bit for multi-user diversity and any further feedback bits for beamforming. We obtain the scaling laws of this compound strategy, showing that it scales as well as any single-beam system with full transmit-CSI.**

***Index Terms*— MIMO, multiuser diversity, opportunistic communication, limited feedback.**

## I. Introduction

In opportunistic beamforming [1] the base-station acquires the channel state information (CSI) from the mobiles (users), then transmits to the user with the best link during each interval. Other methods of multi-user scheduling also exist, for example the base-station may use transmit-CSI is round-robin scheduling for users, by transmitting to each user along the eigen-direction of its channel.

Each of these methods nominally requires feedback of continuous-valued coefficients, which implies unlimited reliable feedback. In [2], [3] the question of opportunistic scheduling with limited feedback has been broached, showing that only one bit of feedback per user is sufficient to capture most of the gain of multi-user diversity for a single antenna transmitter. Also, there exists a good amount of work on quantized beamforming (see [4] for an overview).

In this work we show how multi-user diversity and transmit beamforming can coexist *in a practical scenario of limited feedback.* In other words, we ask the following question: in the presence of limited feedback, what combination of the two methods (opportunistic multi-user vs. deterministic beamforming) should we use, how can this combination be accomplished, and how well does it perform.

We propose a combined strategy where one bit of feedback per user is dedicated to multi-user diversity, and any remaining feedback is used for beamforming. The one-bit multi-user feedback indicates whether each user's channel is above a certain threshold, and the remaining bits choose the beamforming vector using a pre-designed code-book. In the simplest form, this method reduces to opportunistic antenna selection whose sum-rate capacity grows asymptotically as well as opportunistic beamforming with full CSI, as we show in the sequel. By using this as a lower bound to more feedback-intensive methods, we show that all such methods achieve the same capacity growth as full-CSI single-beam systems.

We mention a number of related works in this area. The concept of thresholding for opportunistic communication is due to Gesbert and Alouini [5]. One way of interpreting our work is that we facilitate threshold-based techniques by providing a finite-rate feedback for them. Also several other works in the area of opportunistic communication are related to this work to varying degrees, including [6], [2], [7], [8], [9], [10], [11] . Our emphasis and contribution is on strategies to achieve finite feedback in single-beam systems. Sharif and Hassibi [8] propose a multi-beam system that has a higher capacity growth than single-beam systems. We conjecture that methods similar to the ones developed in this correspondence can be derived for multi-beam systems. In the remainder of the paper, whenever we mention "full-CSI system" we are referring to the single-beam full-CSI system.

## II. System Model

We consider a network of $n$ users each having $N$ antenna for receiving data from the base-station. The base-station has $M$ antennas. For $k^{th}$ user we assume the linear time invariant flat fading model:

$$\mathbf{y}_k(t) = \mathbf{H}_k \cdot \mathbf{x}_k(t) + \mathbf{n}_k(t)$$

where $\mathbf{y}_k(t) \in \mathbb{C}^N$ is the received signal and $x_k(t) \in \mathbb{C}^{M\times 1}$ is the transmitted signal for user $k$ at time $t$. The transmit power is limited by $\rho$, i.e. $\mathbb{E}[\|\mathbf{x}\|^2] \leq \rho$, $\mathbf{n}_k(t)$ is an i.i.d. circularly symmetric complex Gaussian noise distributed according to $\mathbf{n}_k(t) \sim \mathcal{CN}(0, I_N)$ and $\mathbf{H}_k$ is an $N \times M$ channel matrix whose $ij^{th}$ element, $\mathbf{H}_{ij,k}$ represents the channel gain between the $i^{th}$ transmit antenna at the base-station and the $j^{th}$ receive antenna of the $k^{th}$ user. We use the following notation: $\mathbb{E}[\ ]$ refers to expected value of a random variable, $\gamma \approx 0.577$ is the Euler-Mascheroni constant and $\log(.)$ is the natural logarithm. We use $a_n \overset{\circ}{=} b_n$ to denote the asymptotic equivalence of $a_n$ and $b_n$ defined as: $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

The antennas at the base-station are uncorrelated and users experience independent channels. We assume a reliable, limited-rate feedback channel for the CSI. The network is assumed to be under power control so that path loss and shadowing do not come into play.

Manuscript received October 3, 2005; revised May 5, 2006 and January 9, 2007; accepted January 23, 2007. The associate editor coordinating the review of this paper and approving it for publication was D. Gesbert. This work was presented in part in the Asilomar Conference on Signals, Systems, and Computers, 2005.

S. Sanayei is with ArrayComm LLC, San Jose, CA, USA (e-mail: sanayei@ieee.org).

A. Nosratinia is with the University of Texas at Dallas Richardson, TX 75083-0688, USA (e-mail: aria@utdallas.edu).

Digital Object Identifier 10.1109/TWC.2007.05763.

## III. Scheduling with Limited Feedback of CSI

When the base-station has one antenna ($M = 1$), Sanayei and Nosratinia [3] proposed a downlink scheduling algorithm with only one bit of feedback per user, as follows: The base-station sets a threshold $\alpha$ for all users. Each user will send a "1" (eligible user) to the base-station if their channel gain exceeds the threshold, otherwise a "0" is sent. The base-station selects randomly from among eligible users for data transmission. If all the feedback bits received by the base-station are zero, then no signal is transmitted in that interval.[1] For a single-beam system, it was shown [3], that the 1-bit algorithm achieves the same capacity growth as full-CSI feedback subject to judicious choice of the threshold. It was also shown that the 1-bit scheduling actually improves fairness over a full-CSI feedback. The interested reader is referred to [3] for details.

Fig. 1 compares the sum-rate capacity of the 1-bit scheduling and full CSI scheduling, suggesting that there is not much gain in spending more than one bit for quantizing channel gains. *Therefore it is reasonable to use any extra feedback, over and above one bit, for other purposes*. Another way we can use transmit-side CSI is beamforming, and we propose that any excess channel state information, over and above one bit, can be used to exploit beamforming gain.

We assume the feedback rate is limited to $L$ bits per channel. A beamforming code-book $\mathcal{U} = \{u_1, \cdots, u_{2^L}\}$ is shared by all users and the base station. Each user picks the beamformer that leads to the highest gain, i.e.

$$\widehat{u}_k = \arg\max_{u \in \mathcal{U}} \|\mathbf{H}_k u\|^2$$

then it compares the corresponding channel gain $\eta_k = \|\mathbf{H}_k \widehat{u}_k\|^2$ to the threshold value $\alpha$ advertised by the base-station. If the channel gain is above the threshold, the user sends its $L$-bit feedback information to the base-station, otherwise it does not transmit any feedback information. Thus the reception of $L$ bits from user $k$ by the base-station indicates that

1) The user $k$ is *eligible* for transmission
2) The base-station should use the beamforming vector $\widehat{u}_k \in \mathcal{U}$ for transmission to user $k$.

For scheduling, the base-station randomly selects one of the eligible users and when there is no eligible user in the network, it does not transmit to any user.

## IV. Sum-Rate of Opportunistic Beamforming with Limited Feedback

The performance of any opportunistic beamforming system is bounded above by the full-CSI performance, and bounded below by antenna selection (assuming feedback rate no less than $\log\lceil M \rceil$). We shall show that opportunistic antenna selection has capacity growth $\log\log n$, where $n$ is the number of users, i.e., it has the same capacity growth as the full-CSI case. Therefore a sandwich argument shows that *any* opportunistic beamforming method can attain capacity growth $\log\log n$ with

[1]When all feedback bits are "0", the base-station can also randomly pick a user for transmission to avoid waste, however, in the asymptote of large number of users this has vanishing advantage.

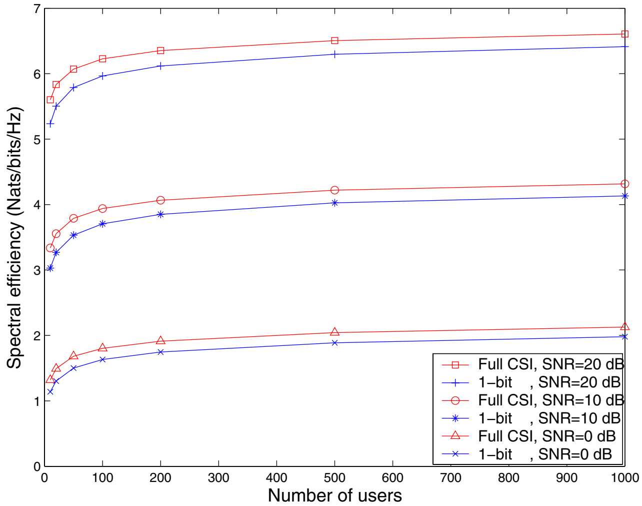


Fig. 1. Comparison of sum-rate capacity for limited and full CSI feedback scheduling for different values of SNR ($M = 1$).

the number of users. The remainder of this section is dedicated to demonstrating the key fact in the argument above, namely to show that opportunistic antenna selection attains $\log\log n$ growth.

When the beamforming codebook is of size $M$ (hence $L = \lceil \log_2 M \rceil$) then the best choice for the code-book is to take $u_l$'s as columns of identity matrix of size $M$,

$$\mathcal{U} = \left\{ \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

In this case, beamforming is equivalent to antenna selection in the base-station (since only one antenna at a time is active).

Let $\mathbf{h}_{i,k}$ denote the $i^{th}$ column of the channel matrix $\mathbf{H}_k$. The user $k$ finds the column of its channel matrix with the maximum norm

$$\widehat{i}_k = \arg \max_{1 \le i \le M} \|\mathbf{h}_{i,k}\|^2$$

and then compares $\|\mathbf{h}_{\widehat{i}_k,k}\|^2$ with the threshold value $\alpha$ advertised by the base-station. If $\|\mathbf{h}_{\widehat{i}_k,k}\|^2 > \alpha$, then the antenna index $\widehat{i}_k$ is transmitted to the base station, otherwise, no feedback is sent. Note that $\widehat{i}_k$ indicates the best transmit antenna for downlink transmission to the user $k$. Upon receipt of this information from eligible users, the base station randomly chooses one user for transmission from among all users whose feedback information has been successfully received and for that user, it uses the corresponding antenna determined by the feedback received from that user.

### A. Sum-Rate Capacity

The equivalent channel gain for the user $j$ is

$$\eta_j = \max_{1 \le i \le M} \|\mathbf{h}_{i,j}\|^2.$$

Under the Rayleigh fading assumption, each channel gain $\|\mathbf{h}_{i,k}\|^2$ is $\chi^2$ distributed with $2N$ degrees of freedom, thus the CDF of $\eta_j$ is

$$F(x) = (1 - e_N(x)e^{-x})^M , \quad (1)$$

where $e_N(x) = \sum_{l=0}^{N-1} \frac{x^l}{l!}$. Let $p = \Pr[\eta_k > \alpha]$. Since the channel gains are all mutually independent, the probability of having $k$ users above the threshold obeys a binomial law, i.e.

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

The ergodic capacity upon having $k$ users above the threshold is:

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

where $C_i = \int_0^\infty \log(1+\rho x) dF_i(x)$ and $F_i(x)$ is the CDF of the $i^{th}$ highest equivalent channel gain. In other words if $\{X_1, \ldots, X_n\}$ is a permutation of $\{\eta_1, \ldots, \eta_n\}$ such that $0 \le X_n \le \ldots \le X_1$, then $F_i(x) = Pr[X_i < x]$. When the channel gains are i.i.d., it can be shown that [12]:

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

Thus the sum-rate capacity of the network with limited feedback can be formulated as:

$$C_{LF} = \sum_{k=1}^{n} p_k \overline{C}_k \quad (5)$$

Thus $C_{LF}$ can be characterized as follows:

$$\begin{aligned} C_{LF} &= \sum_{i=1}^{n} \pi_i \int_0^\infty \log(1+\rho x) dF_i \\ &= \int_0^\infty \log(1+\rho x) dF_\pi \end{aligned} \quad (6)$$

where $F_\pi = \sum_{i=1}^{n} \pi_i F_i$ is a mixture probability measure of all order statistics of the family with parent CDF $F(.)$ and $\{\pi_i\}_{i=0}^n$ is a discrete probability measure defines as

$$\pi_i = \frac{1}{np}\sum_{k=i}^{n} p_k, \quad i = 1, \ldots, n. \quad (7)$$

The exchange of summation and integral leading to Eq. (6) is due to Fubini's theorem, since $F(\cdot)$ and therefore $F_\pi$ have exponential tails, so $\log(1+\rho)$ is absolutely integrable with respect to $F_\pi$.

*1) Optimal Threshold:* The sum-rate capacity is a function of $\rho$, $p$ and $n$. The relation between the threshold $\alpha$ and the probability $p$ is given by

$$\alpha = F^{-1}(1-p). \quad (8)$$

The inverse function of the CDF given by (1) in general can not be explicitly calculated. When users have only one receive antenna ($N = 1$), however,

$$\alpha = -\log\left(1 - (1-p)^{\frac{1}{M}}\right). \quad (9)$$

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

$$\sum_{k=1}^{n} (k - np) p_k \overline{C}_k = 0. \quad (10)$$

A closed form solution to this equation is in general not tractable. However, a numerical solution is possible with $O(n)$ complexity.

### B. Asymptotic Analysis

For simplicity we first assume each user has one antenna ($N = 1$). Then

$$C_{Full\_CSI} = \mathbb{E}[\log(1 + \rho \max_{1\le k\le n} \|\mathbf{h}_k\|^2)] \quad (11)$$

where $\mathbf{h}_k$ is the $1 \times M$ channel vector of user $k$.

*Theorem 1:* The sum-rate capacity of coherent opportunistic beamforming with full CSI available at the base-station scales as

$$C_{Full\_CSI} \stackrel{\circ}{=} \log\log n + \log \rho.$$

**Proof:** The random variable $Y_k = \|\mathbf{h}_k\|^2$ is distributed according to $\chi^2_{2M}$. Using classical results in extreme value theory, it is shown in [13], [14] that the mean and the variance of $Z_n = \max_{1\le k\le n} Y_k$ have the following asymptotic behavior

$$\mu_n = \mathbb{E}[Z_n] \stackrel{\circ}{=} \log n + \log\left(\frac{n^{M-1}}{(M-1)!}\right) + \gamma \quad (12)$$

$$\sigma_n^2 = \mathbb{E}\left[(Z_n - \mathbb{E}[Z_n])^2\right] \stackrel{\circ}{=} \frac{\pi^2}{6} \quad (13)$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Therefore $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$ as $n \to \infty$ thus $Z_k$ satisfies the condition in Theorem 4 (see Appendix) and we have

$$\begin{aligned} C_{Full\_CSI} &= \mathbb{E}[\log(1+\rho Z_n)] \\ &\stackrel{\circ}{=} \log(1+\rho\mathbb{E}[Z_n]) \\ &\stackrel{\circ}{=} \log\left(1 + \rho(\log n + \log\left(\frac{n^{M-1}}{(M-1)!}\right) + \gamma)\right) \\ &\stackrel{\circ}{=} \log\log n + \log\rho. \end{aligned} \quad (14)$$

■

*Theorem 2:* The sum-rate capacity of opportunistic transmit antenna selection scales the same as scheduling with full CSI (coherent beamforming), i.e.

$$\lim_{n\to\infty} \frac{C_{LF}}{C_{Full\_CSI}} = 1.$$

**Proof:** From Equation (6) we have $C_{LF} = \int_0^\infty \log(1+\rho x) dF_\pi$ where

$$F_\pi = \sum_{k=1}^{n} \pi_k F_k$$

and $F_k$'s are the probability measures associated with order statistics with the parent distribution $F(x) = (1 - e^{-x})^M$. If $X$ is a random variable distributed according to $F(\cdot)$, we define the function $g(\cdot)$ such that $X = g(Y)$ where $Y$ is an exponential random variable. We have

$$\begin{aligned} F(x) &= \Pr[X < x] = \Pr[g(Y) < x] \\ &= \Pr[Y < g^{-1}(x)] = 1 - e^{-g^{-1}(x)} \end{aligned}$$

thus

$$y = g^{-1}(x) = -\log(1 - F(x)). \tag{15}$$

We have $y' = \frac{f(x)}{1-F(x)} > 0$ thus $g^{-1}(x)$ and hence $g(x)$ are strictly increasing functions, which means they preserve order. Therefore if $0 < X_n \le \cdots \le X_1$ are the order statistics with parent distribution, $F(\cdot)$, $0 < Y_n \le \cdots \le Y_1$ with $Y_k = g^{-1}(X_k)$ are the order statistics of the exponential distribution.

$$C_{LF} = \sum_{k=1}^{n} \pi_k \mathbb{E}[\log(1 + \rho X_k)] = \sum_{k=1}^{n} \pi_k \mathbb{E}[\log(1 + \rho g(Y_k)].$$

This can be written as

$$C_{LF} = \int_0^\infty \log(1 + \rho \cdot g(x))\, d\nu_\pi(x)$$

where $\nu_\pi = \sum_{k=1}^n \nu_k$ is the mixture probability distribution of all order statistics of the exponential distribution. $\nu_k(x)$ is the CDF of $Y_k$ and can be calculated as follows

$$\nu_k(x) = \sum_{l=0}^{k-1} \binom{n}{l} e^{-lx}(1 - e^{-x})^{n-l}$$

In the proof of Theorem 1, we showed that $\frac{\mu_\pi(\nu)}{\mu_1(\nu)}$ and $\frac{\sigma_\pi(\nu)}{\mu_\pi(\nu)} \to 0$, as $n \to \infty$ where

$$\mu_1(\nu) = \int_0^\infty x d\nu_1(x) = \mathbb{E}[Y_1] = \sum_{k=1}^{n} \frac{1}{k}$$

$$\mu_\pi(\nu) = \int_0^\infty x d\nu_\pi(x) = \sum_{k=1}^{n} \pi_k \mathbb{E}[Y_k]$$

and

$$\sigma_1^2(\nu) = \int_0^\infty (x - \mu_1(\nu))^2 d\nu_1(x) = \sum_{k=1}^{n} \frac{1}{k^2}$$

$$\sigma_\pi^2(\nu) = \int_0^\infty (x - \mu_\pi(\nu))^2 d\nu_\pi(x).$$

*Definition 1:* We define $\mathcal{G}$ as the set of measurable functions $g : \mathbb{R}^+ \mapsto \mathbb{R}^+$ that are both increasing and concave with $g(0) = 0$

$$\mathcal{G} = \{g : \mathbb{R}^+ \mapsto \mathbb{R}^+ | \forall x > 0,\ g'(x) > 0,\ g''(x) < 0,\ g(0) = 0\}.$$

Now we show that the function $g(\cdot)$ belongs to $\mathcal{G}$, thus we can apply Theorem 7 to show that $C_{LF}$ scales the same as $C_{Full\_CSI}$. Recall $g(\cdot)$ is strictly increasing and $g^{-1}(0) = -\log(1 - F(0)) = 0$ hence $g(0) = 0$. To prove that $g(\cdot)$ is concave, it is sufficient to show $y = g^{-1}(x) = -\log(1-(1-e^{-x})^M)$ is convex. We have $1 - e^{-y} = (1 - e^{-x})^M$ hence

$$y' e^{-y} = M e^{-x}(1 - e^{-x})^{M-1}$$

$$y'' e^{-y} - (y')^2 e^{-y} = M e^{-x}(1 - e^{-x})^{M-2}(M e^{-x} - 1)$$

after some algebra we get

$$\begin{aligned} \frac{y''}{y'} e^{-y}(1 - e^{-x}) &= M e^{-x} - (1 - (1 - e^{-x})^M) \\ &= M e^{-x} - e^{-x} \sum_{i=0}^{M-1} (1 - e^{-x})^i \\ &= e^{-x} \sum_{i=0}^{M-1} (1 - (1 - e^{-x})^i) > 0. \end{aligned}$$

The latter inequality means that $y'' > 0$ for all $x > 0$, hence $g^{-1}$ is convex, therefore $g \in \mathcal{G}$. Furthermore, we prove $\lim_{x\to\infty} \frac{g^{-1}(x)}{x} = 1$,

$$\begin{aligned} \lim_{x\to\infty} \frac{g^{-1}(x)}{x} &= \lim_{x\to\infty} -\frac{\log(1 - (1 - e^{-x})^M)}{x} \\ &= \lim_{x\to\infty} \frac{-\log\left(1 - (1 - M e^{-x} + \cdots + (-1)^M e^{-Mx})\right)}{x} \\ &= \lim_{x\to\infty} 1 + \frac{\log\left(M - \binom{M}{2} e^{-x} + \cdots + (-1)^M e^{-(M-1)x}\right)}{x} \\ &= 1. \end{aligned}$$

We also have $\lim_{x\to\infty} g^{-1}(x) = \lim_{x\to\infty} g(x) = +\infty$, let $y = g(x)$, then

$$\lim_{x\to\infty} \frac{g(x)}{x} = \lim_{y\to\infty} \frac{y}{g^{-1}(y)} = 1$$

Therefore $g(\cdot)$ satisfies all the conditions of Theorem 7, hence

$$\begin{aligned} C_{LF} &= \int_0^\infty \log(1 + \rho x) dF_\pi(x) \\ &= \int_0^\infty \log(1 + \rho g(x)) d\nu_\pi(x) \\ &\overset{\circ}{=} \log(1 + \rho g(\mu_\pi(\nu))) \\ &\overset{\circ}{=} \log(1 + \rho \mu_\pi(\nu)) \\ &\overset{\circ}{=} \log(1 + \rho \mu_1(\nu)) \\ &\overset{\circ}{=} \log(1 + \rho \log n) \\ &\overset{\circ}{=} \log\log n + \log \rho. \end{aligned}$$

Thus from Theorem 1 we conclude

$$C_{LF} \overset{\circ}{=} C_{full\_CSI}$$

■

Figure 2 shows a comparison of the throughput of various methods mentioned in this paper. The best performance is that of coherent opportunistic beamforming (CO), where the multiple-antenna transmitter has full knowledge of the (vector) channel of all users, and in each time interval beamforms towards the best user. This method requires generous amounts of feedback.

With a slight loss of throughput, one may use antenna selection, where instead of full beamforming, the transmitter selects one of the antennas. In other words, with the knowledge of user (vector) channels, the transmitter picks the best transmit antenna and user among all possible such pairs.

Opportunistic beamforming (OB) as suggested by Viswanath et al [1] has performance below that of

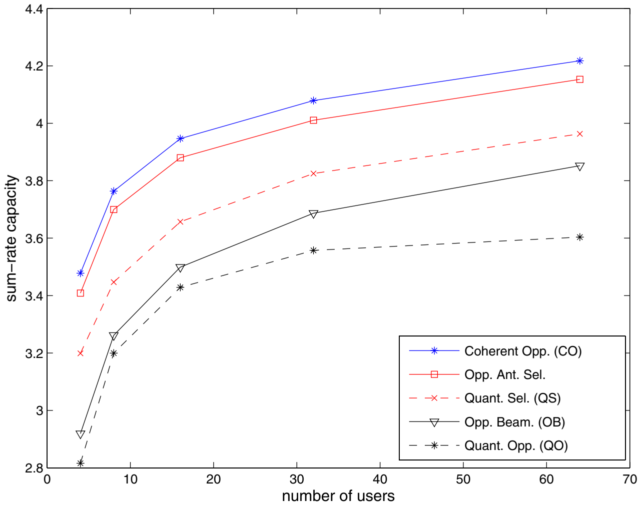


Fig. 2. Comparison of the throughput of various methods at SNR=10dB.

opportunistic antenna selection. Finally, each of the last two cases can be quantized, resulting again in some loss of throughput. The quantized methods shown in Figure 2 are simulated with one bit per user, i.e., with two transmit antennas at the base station, where for transmission to each user the better transmit antenna is selected.

## V. Discussion and Conclusion

In this paper we investigate communication in the presence of strictly limited feedback in the downlink of multi-antenna, multi-user systems. We propose and investigate a method where one bit of feedback is used for multi-user diversity and the remainder of feedback bits are used for array gain, showing that it has growth rate as good as single-beam opportunistic beamforming with unlimited feedback.

It is possible to reduce the feedback rate even further through MAC layer mechanisms. For example, we may allow the users to send their CSI feedback not in dedicated channels, but jointly in a contention-based channel. In this case, the signaling can be designed such that any users whose downlink channel is below the prescribed threshold will stay silent, and only transmit-eligible users will send feedback. This will reduce the load on the feedback channel. A similar approach applies to a CDMA feedback channel, where the users below the threshold stay silent, thus reducing the load on the interference-limited feedback channel and achieving better efficiencies.

## Appendix
## Asymptotic Tightness of Jensen's Inequality

Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a measurable concave function and let $X$ be a random variable, then

$$\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$$

with equality if and only if the function $g(\cdot)$ is an affine function or the probability measure is trivial. This result is known as Jensen's inequality. We wish to investigate when Jensen's inequality is tight for a family of random variables $\{X_n\}_{n=1}^{\infty}$.

*Lemma 1:* For all $x, y \geq 0$ we have

$$|\log(1+\rho x) - \log(1+\rho y)| \leq \log(1+\rho|x-y|).$$

**Proof:** Since $g(x) = \log(1+\rho x)$ is an increasing function, without loss of generality we can assume $x \geq y \geq 0$. Therefore we have $\rho^2 y(x-y) \geq 0$. This inequality can be re-written as $\frac{1+\rho x}{1+\rho y} \leq 1+\rho(x-y)$, and by taking the logarithm of both sides we arrive at the desired inequality. ■

*Definition 2:* The family of random variables $\{X_n\}$ is said to be uniformly integrable if

$$\lim_{c\to\infty} \limsup_n \int_c^{\infty} x \; dF_{|X_n|}(x) = 0.$$

*Theorem 3:* [15] 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 $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

*Lemma 2:* If $\mathbb{E}[|X_n|] < \infty$ for all $n$, then the random variable $X_n$ 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

$$c(1 - F_{|X_n|}(c)) \leq \int_c^{\infty} x \; dF_{|X_n|}(x) \leq \mathbb{E}[|X_n|] < \infty$$

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

$$\begin{aligned} 0 &\leq \int_c^{\infty} x \; dF_{|X_n|}(x) \\ &= -x(1-F_{|X_n|}(x))\,|_c^{\infty} + \int_c^{\infty} \Pr[|X_n| > x] \; dx \\ &= -c(1-F_{|X_n|}(c)) + \int_c^{\infty} \Pr[|X_n| > x] \; dx \\ &\leq \int_c^{\infty} \Pr[|X_n| > x] \; dx \end{aligned}$$

and this proves the lemma. ■

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

**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 $n > N_2$, $\Pr[|X_n| > \frac{\epsilon}{\delta_1}] < \delta_2$, thus for all $n > N = \min\{N_1, N_2\}$,

$$Pr[|a_n X_n| > \epsilon] = \Pr[|X_n| > \frac{\epsilon}{|a_n|}] \leq \Pr[|X_n| > \frac{\epsilon}{\delta_1}] \leq \delta_2$$

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

Now we prove the following theorems for asymptotic tightness of Jensen's inequality,

*Theorem 4:* Let $\{X_n\}$ be a family of positive i.i.d. random variable with finite mean $\mu_n$ and variance $\sigma_n^2$, also $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$ as $n \to \infty$, then for all $\rho > 0$ we have

$$\frac{\mathbb{E}[\log(1+\rho X_n)]}{\log(1+\rho\mathbb{E}[X_n])} \longrightarrow 1. \tag{16}$$

**Proof:** Using Chebyshev's inequality for all $\epsilon > 0$ we have:

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

hence $\frac{1+\rho X_n}{1+\rho\mu_n} \xrightarrow{i.p.} 1$ . Using the *continuous mapping theorem*, we have

$$\log\left(\frac{1+\rho X_n}{1+\rho\mu_n}\right) \xrightarrow{i.p.} 0 \ .$$

On the other hand $\mu_n \to \infty$, hence $\frac{1}{\log(1+\rho\mu_n)} \to 0$ and we can invoke Lemma 3 to conclude

$$\frac{\log\left(\frac{1+\rho X_n}{1+\rho\mu_n}\right)}{\log(1+\rho\mu_n)} = \frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1 \xrightarrow{i.p.} 0$$

Thus

$$\frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} \xrightarrow{i.p.} 1.$$

We show that the random variable $Z_n = \frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1$ is uniformly integrable.

$$\begin{aligned}I &= \int_c^\infty \Pr\left[\left|\frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)}-1\right|>t\right]dt \\ &= \int_c^\infty \Pr\left[\frac{|\log(1+\rho X_n)-\log(1+\rho\log\mu_n)|}{\log(1+\rho\mu_n)}>t\right]dt \\ &\le \int_c^\infty \Pr\left[\frac{\log(1+|X_n-\mu_n|)}{\log(1+\rho\mu_n)}>t\right]dt \qquad \text{by Lemma 1} \\ &= \int_c^\infty \Pr\left[\rho|X_n-\mu_n|>A_n^t-1\right]dt\end{aligned}$$

where $A_n = 1+\rho\mu_n$. Using Chebyshev's inequality we have

$$\begin{aligned}I &\le \int_c^\infty \Pr\left[\rho|X_n-\mu_n|>A_n^t-1\right]dt \\ &\le \rho^2\sigma_n^2\int_c^\infty \frac{dt}{(A_n^t-1)^2}.\end{aligned}$$

We use change of variable $u = A_n^t - 1$, $dt = \frac{du}{\log(A_n)(u+1)}$, $\alpha_n = A_n^c - 1$

$$\begin{aligned}I &\le \rho^2\sigma_n^2\frac{1}{\log(A_n)}\int_{\alpha_n}^\infty \frac{du}{u^2(u+1)} \\ &\le \rho^2\sigma_n^2\frac{1}{\log(A_n)}\int_{\alpha_n}^\infty \frac{du}{u^2} \\ &= \rho^2\left(\frac{\sigma_n}{\mu_n}\right)^2\frac{A_n^2}{\log(A_n)({A_n}^c-1)}.\end{aligned}$$

Since $\mu_n \to \infty$, $A_n \to \infty$ thus for all $c > 2$

$$\frac{A_n^2}{\log(A_n)(A_n^c-1)} \to 0.$$

Also $\frac{\sigma_n}{\mu_n} \to 0$, hence $I \to 0$ and $Z_n = \frac{\log(1+\rho X_n)}{\log(1+\rho\mu_n)} - 1$ is uniformly integrable. Therefore from Theorem 3 we conclude

$$\frac{\mathbb{E}[\log(1+\rho X_n)]}{\log(1+\rho\mu_n)} \underset{n\to\infty}{\longrightarrow} 1 \ .$$

■

We now extend the asymptotic tightness, which was shown for the Shannon function $g(x) = \log(1+\rho x)$, to a larger class of functions. To do so we return to functions in $\mathcal{G}$ (defined in Section IV-B).

*Theorem 5:* $\mathcal{G}$ has the following properties

1) $\mathcal{G}$ is closed under function composition.
2) for every $g \in \mathcal{G}$, $\frac{g(x)}{x}$ is a decreasing function
3) $g \in \mathcal{G}$ is a sub-additive function, hence $\forall x, y > 0$ , $|g(x) - g(y)| < g(|x-y|)$.

**Proof:** For part 1 we note that if $g, f \in \mathcal{G}$ then $g \circ f(x) = g(f(x))$ is defined in $\mathbb{R}^+$. Also $(g \circ f)'(x) = f'(x)g'(f(x)) > 0$ and $(g\circ f)''(x) = f''(x)g'(f(x)) + (f'(x))^2 g''(f(x)) < 0$ therefore $g \circ f \in \mathcal{G}$. For proving part 2, we use the concavity of $g(\cdot)$. For all $x, y > 0$ and $\alpha \in (0,1)$,

$$\alpha g(x) + (1-\alpha)g(y) < g(\alpha x + (1-\alpha)y).$$

Let $y = 0$, then $\alpha g(x) < g(\alpha x)$. This can be written as $\frac{g(x)}{x} < \frac{g(\alpha x)}{\alpha x}$ which means that $g(\cdot)$ is a decreasing function. In part 3, we first prove that $g(\cdot)$ is sub-additive. Using part 2, for all $x, y > 0$ we have $\frac{g(x+y)}{x+y} < \frac{g(x)}{x}$ hence we get

$$\begin{aligned}\frac{x}{x+y}g(x+y) &< g(x)\ , \\ \frac{y}{x+y}g(x+y) &< g(y).\end{aligned}$$

Adding these two inequalities gives $g(x+y) < g(x) + g(y)$, so $g(\cdot)$ is a sub-additive function. Without loss of generality we assume $x \ge y \ge 0$, therefore $g(x) = g(y + (x-y)) < g(y) + g(x-y)$ hence $g(x) - g(y) < g(x-y)$. But $|g(x) - g(y)| = g(x) - g(y)$ because $g(\cdot)$ is increasing. Thus

$$|g(x) - g(y)| < g(|x-y|).$$

■

We now establish the conditions so that Jensen's inequality for any $g \in \mathcal{G}$ is asymptotically tight.

*Theorem 6:* Let $\{X_n\}$ be a family of positive i.i.d. random variable with finite mean $\mu_n$ and variance $\sigma_n^2$, so that $\mu_n \to \infty$ and $\frac{\sigma_n}{\mu_n} \to 0$ as $n \to \infty$, then for every $g \in \mathcal{G}$ satisfying the condition $\lim_{x\to\infty}\frac{g(x)}{x} = K > 0$, we have

$$\frac{\mathbb{E}[g(X_n)]}{g(\mathbb{E}[X_n])} \longrightarrow 1. \qquad (17)$$

**Proof:** As shown in the proof of Theorem 4, $\frac{\sigma_n}{\mu_n} \to 0$ implies $\frac{X_n}{\mu_n} \to 1$. Using the Chebyshev inequality and Theorem 1 (part 3) for every $a > 0$ we have

$$\begin{aligned}\Pr\left[\left|\frac{g(X_n)}{g(\mu_n)}-1\right|>a\right] &= \Pr\left[\left|\frac{g(X_n)-g(\mu_n)}{g(\mu_n)}\right|>a\right] \\ &\le \Pr\left[\frac{g(|X_n-\mu_n|)}{g(\mu_n)}>a\right] \\ &\le \Pr\left[|X_n-\mu_n|>g^{-1}(ag(\mu_n))\right] \\ &\le \left(\frac{\sigma_n}{\mu_n}\right)^2\cdot\left(\frac{\mu_n}{g^{-1}(ag(\mu_n))}\right)^2 \qquad (18)\end{aligned}$$

We have $\mu_n \to \infty$ as $n \to \infty$. Also the condition $\lim_{x\to\infty}\frac{g(x)}{x} = K > 0$, implies that $g(x) \to \infty$ as

$x \to \infty$. $g^{-1}$ is an increasing continuous function therefore $g^{-1}(ag(\mu_n)) \to \infty$, thus

$$\lim_{n\to\infty} \frac{\mu_n}{g^{-1}(ag(\mu_n))} = \lim_{n\to\infty} \frac{g(g^{-1}(ag(\mu_n))}{ag^{-1}(ag(\mu_n))} = \frac{K}{a}$$

Since $\frac{\sigma_n}{\mu_n} \to 0$ from 18 we can conclude that $\frac{g(X_n)}{g(\mu_n)} \xrightarrow{i.p.} 1$. We need to show that the random variable $Z_n = \frac{g(X_n)}{g(\mu_n)} - 1$ is uniformly integrable. Using Lemma 2 and 18 we have

$$\begin{aligned} I_n &= \int_c^\infty \Pr\left[\left|\frac{g(X_n)}{g(\mu_n)} - 1\right| > a\right] da \\ &\le \frac{\sigma_n^2}{\mu_n^2} \int_c^\infty \frac{\mu_n^2 da}{(g^{-1}(ag(\mu_n)))^2}. \end{aligned}$$

We use the change of variable $u = g^{-1}(ag(\mu_n))$, $da = \frac{g'(u)}{g(\mu_n)} du$, $\alpha_n = g^{-1}(cg(\mu_n))$, also we note that $g'(\cdot)$ is a decreasing function thus for $u \in (\alpha_n, \infty)$ we have $g'(u) < g'(\alpha_n)$, hence

$$\begin{aligned} I_n &\le \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \mu_n^2 \int_{\alpha_n}^\infty \frac{g'(u)}{u^2} \frac{du}{g(\mu_n)} \\ &< \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{\mu_n^2 g'(\alpha_n)}{g(\mu_n)} \int_{\alpha_n}^\infty \frac{du}{u^2} \\ &= \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{\mu_n^2 g'(\alpha_n)}{g(\mu_n)} \int_{\alpha_n}^\infty \frac{du}{u^2} \\ &= \left(\frac{\sigma_n}{\mu_n}\right)^2 \cdot \frac{1}{c} \cdot \frac{\mu_n}{g(\mu_n)} \cdot \frac{g(\alpha_n)}{\alpha_n} \cdot g'(\alpha_n). \end{aligned}$$

By L'Hospital's rule, $\lim_{x\to\infty} g'(x) = \lim_{x\to\infty} \frac{g(x)}{x} = K$. Also $\mu_n \to \infty$ implies $\alpha_n \to \infty$, therefore $\lim_{n\to\infty} \frac{\mu_n}{g(\mu_n)} \cdot \frac{g(\alpha_n)}{\alpha_n} \cdot g'(\alpha_n) = K$, hence $\lim_{n\to\infty} I_n = 0$ which means $Z_n = \frac{g(X_n)}{g(\mu_n)} - 1$ is uniformly integrable thus $\mathbb{E}[Z_n] \longrightarrow 0$ ■

**Note:** In the above theorem, if $\limsup \sigma_n < \infty$ then the condition $\lim_{x\to\infty} \frac{g(x)}{x} = K > 0$ can be weakened to $\lim_{x\to\infty} g(x) = +\infty$.

The following theorem combines the results from Theorem 4 and Theorem 6:

*Theorem 7:* Under the assumptions of Theorem 6 we have

$$\frac{\mathbb{E}[\log(1 + \rho g(X_n))]}{\log(1 + \rho g(\mathbb{E}[X_n])} \longrightarrow 1.$$

**Proof:** The proof is straightforward and very similar to the proof of Theorem 4. ■

## REFERENCES


[1] 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, June 2002.

[2] M. Johansson, "Benefits of multiuser diversity with limited feedback," in *Proc. IEEE Intl. Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, Rome, Italy, June 2003.

[3] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with only 1-bit feedback," in *Proc. IEEE Wireless Communication and Networking Conference (WCNC)*, New Orleans, LA, March 2005.

[4] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[5] D. Gesbert and S. Alouini, "How much feedback is multi-user diversity really worth," in *Proc. International Conference on Communications*, Paris, France, June 2004.

[6] T. Tang and R. W. Heath, "Opportunistic feedback for downlink multiuser diversity," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 948–950, Oct. 2005.

[7] E. G. Larsson, "On the combination of spatial diversity and multiuser diversity," *IEEE Commun. Lett.*, vol. 8, no. 8, pp. 517–519, Aug. 2004.

[8] 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.

[9] F. Floren, O. Edfors, and B-A Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme," in *Proc. IEEE Global Telecommunications Conference*, Sanfrancisco,CA, Dec. 2003.

[10] I.-M. Kim, S.-C. Hong, S. S. Ghassemzadeh, and V. Tarokh, "Opportunistic beamforming based on multiple weighting vectors," *IEEE Trans. Wireless Commun.*, vol. 4, no. 11, pp. 2683–2687, Nov. 2005.

[11] D. V. Djonin and V. K. Bhargava, "An upper bound on the throughput of opportunistic transmission in a multiple-access fading channel," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1618–1621, Oct. 2004.

[12] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. John Wiley and Sons, 1992.

[13] S. Sanayei and A. Nosratinia, "Asymptotic capacity analysis of transmit antenna selection," in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, June 2004, p. 241.

[14] S. Sanayei and A. Nosratinia, "Capacity of MIMO channels with antenna selection," to appear in *IEEE Trans. Inf. Theory*.

[15] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.