Multi-Antenna Downlink Channels with Limited Feedback and User Selection

Taesang Yoo, *Student Member, IEEE,* Nihar Jindal, *Member, IEEE,* and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*— We analyze the sum-rate performance of a multi-antenna downlink system carrying more users than transmit antennas, with partial channel knowledge at the transmitter due to finite rate feedback. In order to exploit multiuser diversity, we show that the transmitter must have, in addition to directional information, information regarding the quality of each channel. Such information should reflect both the channel magnitude and the quantization error. Expressions for the SINR distribution and the sum-rate are derived, and tradeoffs between the number of feedback bits, the number of users, and the SNR are observed. In particular, for a target performance, having more users reduces feedback load.**

***Index Terms*— MIMO, quantized feedback, limited feedback, zero-forcing beamforming, multiuser diversity, broadcast channel, scheduling, semi-orthogonal user selection, random beamforming**

I. Introduction

RECENT advances in multiuser downlink communication channels show that in multiple input multiple output (MIMO) systems with M transmit antennas and $K \geq M$ single antenna users, the full multiplexing gain M can be achieved by using space-division multiple access schemes such as dirty-paper coding (DPC) or transmit beamforming [1]–[3]. Moreover, in a large user regime $K \gg M$, the sum-capacity grows like $M \log\log K$ due to multiuser diversity [3]–[5]. Low-complexity schemes based on zero-forcing beamforming (ZFBF) or zero-forcing dirty-paper coding (ZF-DPC) have been proposed that achieve this optimal growth rate [1], [6]–[9]. However, all these results are based on the assumption of perfect channel state information at the transmitter (CSIT), which may not be a practical assumption.

One of the popular models to address the lack of perfect CSIT is to provide the transmitter with imperfect CSI via a rate constrained feedback channel from each user, where each user quantizes its vector channel to one of $N = 2^B$ quantization vectors and feeds back the corresponding index. This feedback is used to capture information regarding only the spatial direction of the channel (referred to as channel direction information, or CDI), and not the channel magnitude. MIMO systems under limited feedback have been studied for single user systems [10]–[13] and recently applied to downlink systems for $K \leq M$ [14]–[16]. For single user systems, it has been shown that only a few feedback bits (roughly on the order of M, the number of transmit antennas) are needed to achieve near perfect-CSIT performance. For downlink channels, however, the feedback load per user must be scaled with both the number of transmit antennas as well as the system SNR in order to achieve near-perfect CSIT performance and the full multiplexing gain [14].

When there are more users than antennas ($K \geq M$), CDI can be used to achieve the full multiplexing gain of the downlink channel, but cannot simultaneously benefit from multiuser diversity, i.e. obtain the double logarithmic growth with K. As we later show, the sum rate with only CDI at the transmitter is bounded as the number of users is taken to be large while all other parameters (feedback load, number of antennas, and SNR) are held constant. In order to scale the sum rate at the optimal $\log\log K$ rate, the transmitter must also have channel quality information (CQI), based on SINR, to exploit selection diversity among users as well as control the effect of quantization error in the CDI. Indeed, the random beamforming (RBF) scheme proposed in [2] uses SINR feedback and a few ($\log_2 M$) additional feedback bits to perform user selection and achieves the asymptotic sum-capacity as $K \to \infty$. However, its performance is generally poor for practical values of K [6].

In this paper, we consider a limited feedback model where each user feeds back B-bit quantized CDI as well as (unquantized) CQI. We propose a low-complexity scheme with a user selection based on a semi-orthogonal user selection (SUS) principle [6], [7], [17], [18] and a ZFBF precoder. When $B = \log_2 M$, our model reduces to the RBF. We characterize the sum-rate performance of our limited feedback model and show how it scales with K.[1] Our analysis reveals tradeoffs between B, K, and SNR, and provides useful design guidelines. Our key findings and results are summarized below:

1) To achieve both multiplexing and multiuser diversity gains, i.e. $M \log\log K$ growth of the sum-rate, both CQI and CDI feedbacks are necessary, and CQI should be the SINR rather than just the channel magnitude. This implies that any quantization should be applied to the effective channel (channel gain divided by noise plus interference) rather than directly to the channel itself.
2) To achieve a constant SNR gap from perfect-CSIT performance, the number of CDI quantization bits (B) and the number of users (K) should scale with SNR (P)

Manuscript received December 31, 2006; revised February 14, 2007. This work was supported in part by the National Science Foundation (NSF) under CCR-0325639-001, LG electronics, and Hitachi Corporation.

Taesang Yoo and Andrea Goldsmith are with Dept. of Electrical Engineering, Stanford University, Stanford, CA 94305 (e-mail: {yoots, andrea}@wsl.stanford.edu).

Nihar Jindal is with Dept. of Electrical & Computer Engineering, University of Minnesota, Minneapolis, MN 55455 (e-mail: nihar@umn.edu).

Digital Object Identifier 10.1109/JSAC.2007.070920.

[1] Although we will extensively use the extreme value theory [19] [20] [21] to derive large K asymptotic behaviors, we will see from numerical results that the asymptotic results carry well to finite K regimes.

as

$$B + \log_2 K = (M-1)\log_2 P + c. \quad (1)$$

Therefore, having more users reduces feedback load.

3) Our limited feedback scheme generalizes the RBF and provides a way to use more CDI feedback bits to improve performance.

The rest of this paper is organized as follows. In Section II, our limited feedback system model is introduced. Section III is devoted to a review of previous results on limited feedback MIMO systems for $K = 1$ and $K \leq M$. In Section IV we consider $K > M$ and analyze the performance under CDI and magnitude-based CQI feedback scheme. Section V contains our main results on the performance of CDI and SINR-based CQI feedback scheme for $K > M$. Finally, we give numerical results in Section VI and conclude in Section VII.

As for notations, we use uppercase boldface letters for matrices and lowercase boldface for vectors. Sets are represented as uppercase calligraphic letters. $\mathbb{E}(\ )$ stands for the expectation operator, and $P(\ )$ is the probability of the given event. $\mathbf{X}^{\mathrm{T}}$ ($\mathbf{x}^{\mathrm{T}}$) stands for the transpose of a matrix $\mathbf{X}$ (vector $\mathbf{x}$), and $\mathbf{X}^*$ ($\mathbf{x}^*$) stands for the conjugate transpose of a matrix $\mathbf{X}$ (vector $\mathbf{x}$). Similarly, $\mathbf{X}^\dagger$ denotes the pseudo-inverse $\mathbf{X}^*(\mathbf{X}\mathbf{X}^*)^{-1}$. $|\mathcal{A}|$ denotes the size of a set $\mathcal{A}$.

II. System Model

Consider a single cell MIMO downlink channel with M transmit antennas at the base and $K \geq M$ users. Our system model is depicted in Figure 1. For simplicity, we assume users have single receive antennas. The results, however, can be extended to multi-antenna users by using techniques developed in [22]. We assume users are homogeneous and experience flat Rayleigh fading. The signal received by a user k can be represented as

$$y_k = \mathbf{h}_k \mathbf{x} + z_k, \quad k = 1, \cdots, K, \quad (2)$$

where $\mathbf{h}_k \in \mathbb{C}^{1\times M}$ is the channel gain vector with zero-mean unit variance i.i.d complex Gaussian entries, $\mathbf{x}$ is the transmitted symbol vector containing information symbols of a selected set of users $\mathcal{S} = \{\pi(1), \cdots, \pi(|\mathcal{S}|)\}$ with an average power constraint $\mathbb{E}\{\|\mathbf{x}\|^2\} = P$, z_k is the additive noise with unit variance, and y_k is the symbol received by user k. The transmit symbol vector $\mathbf{x}$ is related to information symbols $s_i,\ i \in \mathcal{S}$, via linear beamforming $\mathbf{x} = \sum_{i\in\mathcal{S}} \mathbf{w}_i s_i$. Therefore, the received symbol is given by

$$y_k = (\mathbf{h}_k \mathbf{w}_k) s_k + \sum_{j\in\mathcal{S}, j\neq k} (\mathbf{h}_k \mathbf{w}_j) s_j + z_k, \quad k \in \mathcal{S}. \quad (3)$$

The user set $\mathcal{S}$ is chosen to maximize the sum-rate. Below we describe each building block in more detail.

A. Finite rate feedback model for CDI quantization

We assume that each user has perfect knowledge (CSIR) of $\mathbf{h}_k$ and quantizes the direction of its channel $\tilde{\mathbf{h}}_k = \mathbf{h}_k/\|\mathbf{h}_k\|$ to a unit norm vector $\hat{\mathbf{h}}_k$. The quantization is chosen from a codebook of unit norm row vectors of size $N = 2^B$

$$\mathcal{C}_k = \{\mathbf{c}_{k1}, \cdots, \mathbf{c}_{kN}\}, \quad N = 2^B, \quad (4)$$

as $\hat{\mathbf{h}}_k = \mathbf{c}_{kn}$ according to the minimum distance criterion [10]–[12],

$$n = \arg \max_{1\leq j\leq N} \left|\tilde{\mathbf{h}}_k \mathbf{c}_{kj}^*\right|. \quad (5)$$

The codebook $\mathcal{C}_k$ is designed off-line and known to the transmitter and the user k a priori. Each user feeds back only the index n to the transmitter, requiring B feedback bits per user.

B. CQI feedback model

In addition to the CDI, each user feeds back its CQI $g(\mathbf{h}_k)$. We consider two definitions of CQI: one using the channel norm $g(\mathbf{h}_k) = \|\mathbf{h}_k\|^2$ in Section IV, and the other using the SINR $g(\mathbf{h}_k) = \mathrm{SINR}_k$ as the CQI in Section V, among which we show the latter achieves multiuser diversity. We assume the CQI is directly fed back *without quantization*, in order to concentrate on the effect of quantization of CDI. We expect the number of bits for quantizing CQI can be kept relatively small.

C. User selection

Based on $\{g(\mathbf{h}_k)\hat{\mathbf{h}}_k, k = 1, \cdots, K\}$, the transmitter performs user selection and linear beamforming to support up to M out of K users at a time. Since finding the optimal user set that maximizes the sum-rate requires an exhaustive search which is not computationally feasible for moderate to large K, we use a heuristic user selection algorithm based on the semi-orthogonal user selection (SUS) procedure [6], [7]. Specifically, the transmitter selects the first user from the initial user set $\mathcal{A}_0 = \{1, \cdots, K\}$ as

$$\pi(1) = \arg \max_{k\in\mathcal{A}_0} g(\mathbf{h}_k). \quad (6)$$

After selecting i users, the $(i+1)$th user is selected within the user set

$$\mathcal{A}_i = \{1 \leq k \leq K : |\hat{\mathbf{h}}_k \hat{\mathbf{h}}_{\pi(j)}^*| \leq \epsilon,\ 1 \leq j \leq i\} \quad (7)$$

as

$$\pi(i+1) = \arg \max_{k\in\mathcal{A}_i} g(\mathbf{h}_k), \quad (8)$$

where ϵ is a design parameter that dictates the maximum spatial correlation allowed between quantized channels. In this way, the transmitter can choose users that have high channel qualities and are mutually semi-orthogonal in terms of their quantized directions $\hat{\mathbf{h}}_k$. Under perfect CSIT, this user selection method achieves the optimal sum-capacity growth rate $M \log\log K$ at large K and performs quite well for moderate K as well [6], [7].

D. Zero-forcing beamforming (ZFBF)

After a user set $\mathcal{S}$ is selected, the users in $\mathcal{S}$ are supported via ZFBF [1], in which the unit-norm beamforming vectors $\mathbf{w}_i \in \mathcal{C}^{M\times 1}$, $i \in \mathcal{S}$, are chosen to satisfy

$$\hat{\mathbf{h}}_j \mathbf{w}_i = 0, \quad \forall j \neq i, \quad j \in \mathcal{S}. \quad (9)$$

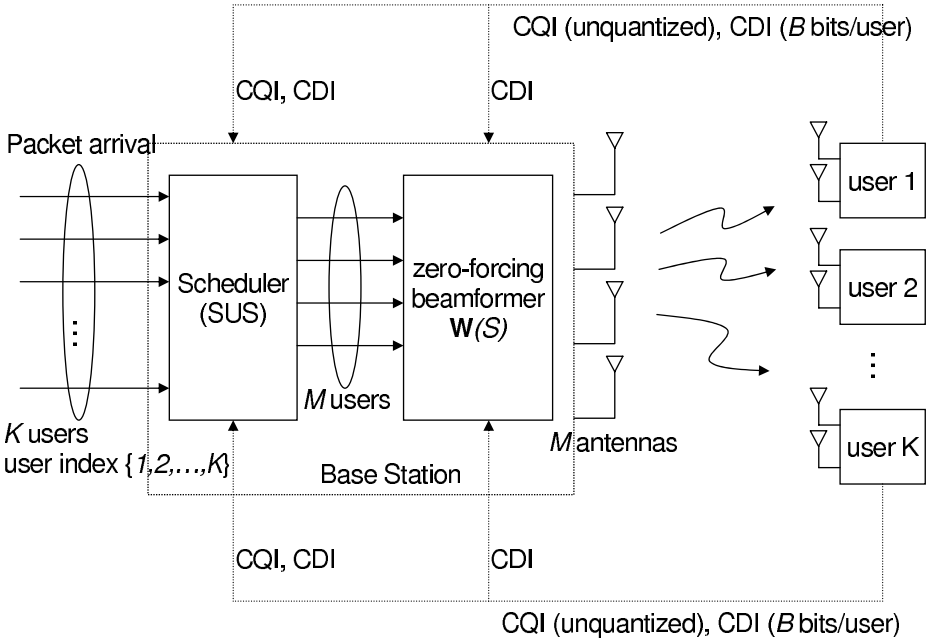

Fig. 1. MIMO downlink system model with CQI and CDI feedbacks and user selection.

Define $\mathbf{H}(\mathcal{S}) = ([\hat{\mathbf{h}}_{\pi(1)}^{\mathrm{T}}, \cdots, \hat{\mathbf{h}}_{\pi(|\mathcal{S}|)}^{\mathrm{T}}]^{\mathrm{T}})$. Then, the beamforming vectors in (9) can be readily determined from the pseudo-inverse

$$\mathbf{W}(\mathcal{S}) = \mathbf{H}(\mathcal{S})^{\dagger} = \mathbf{H}(\mathcal{S})^{*}(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^{*})^{-1}, \qquad (10)$$

with $\mathbf{w}_{\pi(i)}$ obtained by normalizing the ith column of $\mathbf{W}$. It is worthwhile to mention that ZFBF is not the optimal choice[2] among linear beamformers $\mathbf{x} = \sum_{i\in\mathcal{S}} \mathbf{w}_i s_i$. Nevertheless, in this paper we only consider ZFBF for several reasons. Firstly, its analytical simplicity enables us to derive closed-form expressions for performance. Second, at high SNR, ZFBF is asymptotically optimal. Moreover, if the selected users are semi-orthogonal, the gain of any optimal transmission scheme over ZFBF is small, since the channel $\mathbf{H}(\mathcal{S})$ is close to unitary. Lastly, in ZFBF, CDI feedback is sufficient for determining beamforming vectors (as is illustrated in Figure 1), hence CQI is necessary only for the user selection purpose.

E. Quantization codebook model

Although a near-optimal codebook can be designed numerically by minimizing the maximum inner product between any codewords, finding the optimal codebook is difficult except for special cases [11], [12], not to mention any analytical performance characterization based on a codebook designed in such a way. Therefore, instead of explicitly designing a codebook, we resort to a quantization cell approximation used in [12], [13]. The approximation is based on the ideal assumption that each quantization cell is a Voronoi region of a spherical cap with the surface area 2^{-B} of the total surface area of the unit sphere, as is illustrated in Figure 2. Specifically, for a given codebook $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_N\}$, we approximate the actual quantization cell

$$\mathcal{R}_i = \{\tilde{\mathbf{h}} : |\tilde{\mathbf{h}}\mathbf{c}_i^*|^2 \geq |\tilde{\mathbf{h}}\mathbf{c}_j^*|^2, \forall j \neq i\} \qquad (11)$$

as

$$\mathcal{R}_i \approx \{\tilde{\mathbf{h}} : |\tilde{\mathbf{h}}\mathbf{c}_i^*|^2 \geq 1 - \delta\}, \qquad (12)$$

where $\delta = 2^{-\frac{B}{M-1}}$ to give $P\{\mathcal{R}_i\} = 2^{-B}$. From this, the cumulative distribution function (CDF) of $\sin^2\theta_k$ is obtained as

$$F_{\sin^2\theta}(x) = \begin{cases} 2^B x^{M-1}, & 0 \leq x \leq \delta \\ 1, & x \geq \delta. \end{cases} \qquad (13)$$

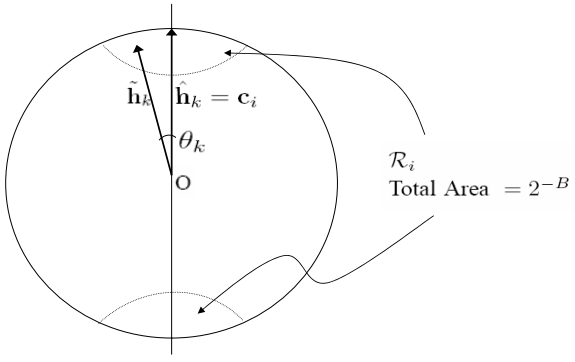

Fig. 2. Quantization cell upper bound (QUB). The figure shows the quantization cell $\mathcal{R}_i$, which is the union of two spherical caps of a total surface area 2^{-B} around the straight line created by the codeword vector $\mathbf{c}_i$. Any channel vector $\tilde{\mathbf{h}}_k$ that falls into this region will be quantized to $\mathbf{c}_i$ according to the QUB model. The area 2^{-B} is normalized such that the entire surface area of the sphere becomes unity.

[2]In general, finding the optimal $\mathbf{w}_i$ that maximizes the sum-rate is difficult, since it is a non-convex optimization problem [23, p. 211].

It is shown in [13] that for *any* quantization codebook $\tilde{\mathcal{C}}$ and its corresponding CDF $F_{\sin^2\tilde{\theta}}$, we have $F_{\sin^2\theta}(x) \geq F_{\sin^2\tilde{\theta}}(x)$. Therefore, the quantization cell approximation yields a performance upper bound, e.g. higher rate, lower outage probability, etc.

To see how tight the upper bound is, in numerical results in Section VI, we evaluate the performance of our proposed system using both the above quantization cell upper bound (QUB) and an actual codebook design based on a random vector quantization (RVQ) scheme [14], [24], [25]. In RVQ, each codeword is randomly and independently generated from M dimensional unit-norm complex Gaussian vectors. Since any reasonably well-designed codebook should perform at least as well as this randomly designed codebook, the RVQ will give a performance lower bound, e.g. lower rate, higher outage probability, etc. As we will see later in Section VI, the difference in the performance between the QUB and the RVQ is small, meaning that analysis based on the quantization cell approximation should serve as a very accurate performance indicator for any well-designed codebook.

We will assume that each unit-norm codeword is isotropically distributed in $\mathbb{C}^{1\times M}$, and that the codewords of different users are independent, i.e. $\mathbf{c}_{k_1 i}$ and $\mathbf{c}_{k_2 j}$ are independent for $k_1 \neq k_2$. These assumptions certainly hold for the RVQ. They are also true if we generate each user's codebook by a random rotation of a basis codebook as follows [15]. First, we generate a basis codebook $\mathcal{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_N\}$. Here, it is useful to employ a matrix notation for the codebook as $\mathbf{C} = [\mathbf{c}_1^T \mathbf{c}_2^T \cdots \mathbf{c}_N^T]^T$. Then, the codebook of the user k is given by $\mathbf{C}_k = \mathbf{C}\mathbf{U}_k$, where $\mathbf{U}_k$ is an M by M random unitary matrix.

III. Review of $K = 1$ and $K = M$

Limited feedback beamforming for single user MIMO systems (or MISO systems since the user has only one antenna) has been studied in [10]–[13]. Note that for a single user system, CQI is not necessary since there is no need of user selection. With a quantized CDI feedback $\hat{\mathbf{h}}_1$, a reasonable transmission strategy is to beamform to $\hat{\mathbf{h}}_1$. This beamforming strategy is indeed optimal under certain conditions and generally very close to optimum [26]. Comparing the rate of this quantized CDI feedback scheme, given by

$$\begin{aligned} R_{\text{CDI}} &= \mathbb{E}\left(\log_2\left(1 + P|\mathbf{h}_1\hat{\mathbf{h}}_1^*|^2\right)\right) \\ &= \mathbb{E}\left(\log_2\left(1 + P\|\mathbf{h}_1\|^2\cos^2\theta_1\right)\right), \end{aligned} \tag{14}$$

where $\theta_1 = \arccos(|\mathbf{h}_1\hat{\mathbf{h}}_1^*|)$ is the angle between $\mathbf{h}_1$ and $\hat{\mathbf{h}}_1$, against the capacity under perfect CSIT

$$R_{\text{CSIT}} = \mathbb{E}\left(\log_2\left(1 + P\|\mathbf{h}_1\|^2\right)\right), \tag{15}$$

we observe that the loss due to quantization appears as an SNR degradation by a factor of $\cos^2\theta_1$. Therefore, as long as θ_1 is not too large, the performance loss is small. Typically, B on the order of M suffices for good performance.[3]

[3]It is shown in [14] that $B = M - 1$ gives a performance within about 3dB of a perfect CSIT system.

Now we look at the case of $K = M$ with a ZFBF transmitter. As was discussed in Subsection II-D, CQI feedback is still not necessary for ZFBF. We begin from (3):

$$y_k = (\mathbf{h}_k\mathbf{w}_k)s_k + \sum_{1\leq j\leq M, j\neq k}(\mathbf{h}_k\mathbf{w}_j)s_j + z_k, \quad 1 \leq k \leq M, \tag{16}$$

which describes the effective channel after the ZFBF. If the CDI was perfect (i.e., $\hat{\mathbf{h}}_k = \tilde{\mathbf{h}}_k = \mathbf{h}_k/\|\mathbf{h}_k\|$), the second term (multiuser interference) would be evaluated to zero (See (9)), and assuming equal power allocation $\rho = P/M$ to the M users, the SINR of the kth user would become

$$\text{SINR}_k = \rho\|\mathbf{h}_k\|^2|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2 = \rho\|\mathbf{h}_k\|^2\beta(1, M-1) \tag{17}$$

where $\beta(1, M-1)$ is a Beta-distributed random variable with parameters $(1, M-1)$ and independent of $\|\mathbf{h}_k\|^2$. The second equality follows since $\mathbf{w}_k$ and $\mathbf{h}_k$ are independent and isotropically distributed in $\mathbb{C}^{1\times M}$. They are independent since $\mathbf{w}_k$ is determined by $\hat{\mathbf{h}}_j$, $1 \leq j \leq M$, $j \neq k$, according to (9), and $\hat{\mathbf{h}}_j$, $j \neq k$, is independent of $\mathbf{h}_k$.

Under quantized CDI, however, the interference term in (16) is not completely eliminated because the beamforming vectors are chosen orthogonal to the quantized channels and not the actual channels. Assuming equal power allocation $\rho = P/M$ to the M users, the SINR of the users are given as

$$\begin{aligned} \text{SINR}_k &= \frac{\rho|\mathbf{h}_k\mathbf{w}_k|^2}{1 + \rho\sum_{j\neq k}|\mathbf{h}_k\mathbf{w}_j|^2} \\ &= \frac{\rho\|\mathbf{h}_k\|^2|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2}{1 + \rho\|\mathbf{h}_k\|^2\sum_{j\neq k}|\tilde{\mathbf{h}}_k\mathbf{w}_j|^2}, \ 1 \leq k \leq M. \end{aligned} \tag{18}$$

Denote as θ_k the angle between $\mathbf{h}_k$ and $\hat{\mathbf{h}}_k$, i.e. $\cos\theta_k = |\tilde{\mathbf{h}}_k\hat{\mathbf{h}}_k^*|$, and decompose $\tilde{\mathbf{h}}_k$ as

$$\tilde{\mathbf{h}}_k = (\cos\theta)\hat{\mathbf{h}}_k + (\sin\theta)\mathbf{g}_k. \tag{19}$$

Applying (9) we can rewrite (18) as

$$\text{SINR}_k = \frac{\rho\|\mathbf{h}_k\|^2|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2}{1 + \rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\sum_{j\neq k}|\mathbf{g}_k\mathbf{w}_j|^2}. \tag{20}$$

Note that both $\mathbf{g}_k$ and $\mathbf{w}_j$ are unit vectors on the $M-1$ dimensional hyperplane orthogonal to $\hat{\mathbf{h}}_k$. Also, since $\mathbf{w}_j$ is solely determined by $\hat{\mathbf{h}}_i$, $i \neq j$, $i \neq k$ within the hyperplane, we conclude that $\mathbf{w}_j$ is isotropic within the hyperplane and independent of $\mathbf{g}_k$. Therefore, (20) becomes [14]

$$\text{SINR}_k = \frac{\rho\|\mathbf{h}_k\|^2\beta(1, M-1)}{1 + \rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\sum_{j\neq k}\beta(1, M-2)}. \tag{21}$$

The random variables $\|\mathbf{h}_k\|^2$, θ_k, and $\beta(1, M-2)$ are all independent. Comparing (21) to (17), we can see that the effect of CDI quantization error is to reduce the SINR by the denominator in (21). Under the quantized cell approximation,

the expected SINR degradation can be evaluated as

$$\begin{aligned}\Delta &= \mathbb{E}\left\{1+\rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\sum_{j\neq k}\beta(1,M-2)\right\}\\ &= 1+\rho\mathbb{E}\left\{\|\mathbf{h}_k\|^2\right\}\mathbb{E}\left\{\sin^2\theta_k\right\}(M-1)\mathbb{E}\left\{\beta(1,M-2)\right\}\\ &= 1+\rho M\left(\int_0^1 x\ dF_{\sin^2\theta}(x)\right)(M-1)\frac{1}{M-1}\\ &= 1+P\left(\frac{M-1}{M}\right)2^{-\frac{B}{M-1}}. \end{aligned} \tag{22}$$

Note that (22) increases linearly in P. Hence, as was noted in [14], for a fixed B, the system becomes interference limited as $P\to\infty$. To maintain the gap Δ constant over P, B needs to be scaled as

$$B=(M-1)\log_2 P+c \tag{23}$$

for some constant c. The same result has been derived in [14] under RVQ. Therefore, the feedback requirement grows linearly with SNR in dB.

IV. $K>M$: CDI and Magnitude Feedback

In this section we consider the case of $K>M$ and analyze the performance of a naive CQI feedback scheme where each user feeds back its quantized CDI $\hat{\mathbf{h}}_k$ as well as its channel magnitude $g(\mathbf{h}_k)=\|\mathbf{h}_k\|^2$.

We start from the SINR expression in (20):

$$\text{SINR}_k=\frac{\rho\|\mathbf{h}_k\|^2|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2}{1+\rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\sum_{j\neq k}|\mathbf{g}_k\mathbf{w}_j|^2},\ k\in\mathcal{S}. \tag{24}$$

For given $\mathbf{h}_k$ and $\hat{\mathbf{h}}_k$ the expected SINR at user k is given by

$$\begin{aligned}\mathbb{E}(\text{SINR}_k)&=\mathbb{E}\left(\frac{\rho\|\mathbf{h}_k\|^2|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2}{1+\rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\sum_{j\neq k}|\mathbf{g}_k\mathbf{w}_j|^2}\right)\\ &\geq\frac{\rho\|\mathbf{h}_k\|^2\mathbb{E}\left(|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2\right)}{1+\rho\|\mathbf{h}_k\|^2(\sin^2\theta_k)\mathbb{E}\left(\sum_{j\neq k}|\mathbf{g}_k\mathbf{w}_j|^2\right)}\\ &=\frac{\rho\|\mathbf{h}_k\|^2\mathbb{E}\left(|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2\right)}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}, \end{aligned} \tag{25}$$

where the inequality follows from Jensen's inequality. Unlike the previous case of $K=M$, $\mathbf{w}_k$ and $\mathbf{h}_k$ are no longer independent, since the channels of selected users are related through the user selection process. Note that the quantized channel $\hat{\mathbf{h}}_k$ is semi-orthogonal to $\hat{\mathbf{h}}_j$ for $j\neq k$. Since the beamforming direction $\mathbf{w}_k$ is orthogonal to $\hat{\mathbf{h}}_j$, $j\neq k$, this means that the two vectors $\hat{\mathbf{h}}_k$ and $\mathbf{w}_k$ are closely aligned. Specifically, denote the angle between the two vectors as ϕ_k, as is illustrated in Figure 3. Then, it has been shown that [7]

$$\cos^2\phi_k=|\hat{\mathbf{h}}_k\mathbf{w}_k|^2>\frac{(1+\epsilon)(1-(M-1)\epsilon)}{(1-(M-2)\epsilon)}\triangleq\cos^2\phi, \tag{26}$$

where ϵ is the semi-orthogonality threshold defined in II-C. Denote the angle between $\tilde{\mathbf{h}}_k$ and $\mathbf{w}_k$ as φ_k. Then, we have $\varphi_k\leq\theta_k+\phi_k$, and assuming $\theta_k+\phi\leq\frac{\pi}{2}$, the squared inner product $|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2$ is lower bounded as $|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2=\cos^2\varphi_k\geq\cos^2(\theta_k+\phi_k)\geq\cos^2(\theta_k+\phi)$. Therefore,

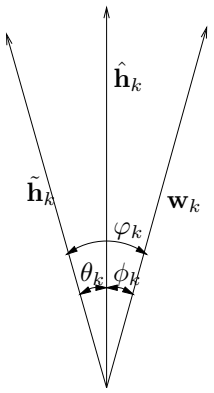

Fig. 3. Normalized channel $\tilde{\mathbf{h}}_k$, its quantization $\hat{\mathbf{h}}_k$, and beamforming direction $\mathbf{w}_k$.

$$\begin{aligned}\mathbb{E}(\text{SINR}_k)&\geq\frac{\rho\|\mathbf{h}_k\|^2\mathbb{E}\left(|\tilde{\mathbf{h}}_k\mathbf{w}_k|^2\right)}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}\\ &\geq\frac{\rho\|\mathbf{h}_k\|^2\cos^2(\theta_k+\phi)}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}\triangleq\gamma_k(\phi), \end{aligned} \tag{27}$$

where ϕ is a constant given by $\phi=\cos^{-1}\sqrt{\frac{1-(M-1)\epsilon}{1-(M-2)\epsilon}(1+\epsilon)}$.

Remark 1: The expectation is conditioned on $\mathbf{h}_k$ and $\hat{\mathbf{h}}_k$, which are known to user k, and is taken over beamforming vectors $\{\mathbf{w}_j: j\in\mathcal{S}, j\neq k\}$, which the user k knows lie in the subspace orthogonal to $\hat{\mathbf{h}}_k$. Therefore, (27) can be interpreted as a lower bound to the SINR_k that user k can expect based on its local information.

We see that in order to maximize the SINR lower bound, the transmitter should try to select mutually semi-orthogonal users (small ϵ). The inequality (27) is tight when ϵ is small, and when $\epsilon=0$, it can be shown that the SINR itself becomes

$$\text{SINR}_k=\frac{\rho\|\mathbf{h}_k\|^2\cos^2\theta_k}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}=\gamma_k(0)\triangleq\gamma_k. \tag{28}$$

Henceforth, for ease of analysis, we use γ_k which approximates SINR_k when ϵ is small.

To obtain an upper bound on the sum-rate, we upper-bound γ_k as

$$\gamma_k=\frac{\rho\|\mathbf{h}_k\|^2\cos^2\theta_k}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}\leq\frac{\cos^2\theta_k}{\sin^2\theta_k}\triangleq\tilde{\gamma}_k, \tag{29}$$

whose distribution depends on the quantization codebook design. We use the quantization cell upper-bound (QUB) explained in Subsection II-E. From (13) the CDF of $\tilde{\gamma}_k$ is derived as

$$\begin{aligned}F_{\tilde{\gamma}_k}(x)&=P\left(\frac{\cos^2\theta_k}{\sin^2\theta_k}\leq x\right)\\ &=1-F_{\sin^2\theta_k}\left(\frac{1}{x+1}\right)\\ &=\begin{cases}1-\frac{2^B}{(x+1)^{M-1}}, & x\geq\frac{1}{\delta}-1\\ 0, & 0\leq x\leq\frac{1}{\delta}-1.\end{cases}\end{aligned} \tag{30}$$

The probability density function (PDF) of $\tilde{\gamma}_k$ is given by

$$f_{\tilde{\gamma}_k}(x) = \begin{cases} \frac{2^B(M-1)}{(x+1)^M}, & x \geq \frac{1}{\delta} - 1 \\ 0, & 0 \leq x < \frac{1}{\delta} - 1. \end{cases} \tag{31}$$

Finally, the expected sum-rate for a selected semi-orthogonal user set $\mathcal{S}$, assuming $\phi \approx 0$, is given by

$$\begin{aligned} \mathbb{E}\{R\} &\approx \mathbb{E}\left\{\sum_{i\in\mathcal{S}} \log_2(1+\gamma_i)\right\} \\ &\leq \mathbb{E}\left\{\sum_{i\in\mathcal{S}} \log_2(1+\tilde{\gamma}_i)\right\} \qquad (32) \\ &= \sum_{i\in\mathcal{S}} \int_0^\infty f_{\tilde{\gamma}_i}(x)\ \log_2(1+x)\ dx \\ &= M \int_{\frac{1}{\delta}-1}^\infty \frac{2^B(M-1)}{(x+1)^M} \log_2(1+x)\ dx \\ &= \frac{M}{M-1}(B + \log_2 e), \qquad (33) \end{aligned}$$

which is independent of P and K. This means that, for a fixed B, the system not only becomes interference limited as either P or K increases, but also does not benefit from multiuser diversity even with $\|\mathbf{h}_k\|^2$ feedback.[4] This is because the SINR is essentially limited by the CDI quantization error θ_k, of which the transmitter has no knowledge. Note that since $\lim_{P\to\infty} \gamma_i = \tilde{\gamma}_i$, the inequality in (32) becomes an equality as $P \to \infty$. Therefore, the upper bound (33) is tight at high SNR. Note also that at high SNR the upper bound (33) can be approached even *without* any magnitude feedback, since $\lim_{P\to\infty} \gamma_i = \tilde{\gamma}_i$ is independent of the magnitude $\|\mathbf{h}_i\|^2$. Therefore, we conclude that CQI should not be solely based on the channel magnitude, especially at high SNR. Instead, a good CQI should consider both the channel magnitude ($\|\mathbf{h}_k\|^2$) and the CDI quantization error (θ_k), which motivates the use of $\gamma_k(\phi)$ as the CQI in the next section.

V. $K > M$: CDI AND SINR FEEDBACK

In this section we analyze the sum-rate of the finite-rate feedback scheme with SINR-based CQI. Note that the exact SINR in (18) is unknown at either the transmitter or receiver. Therefore, we propose the use of $g(\mathbf{h}_k) = \gamma_k(\phi)$ in (27) which, as noted in Remark 1, the user k can calculate based on $\mathbf{h}_k$, θ_k, and a given parameter ϕ. To simplify analysis, we assume the feedback takes the form $g(\mathbf{h}_k) = \gamma_k(0) = \gamma_k$. In the following lemmas we derive the distribution of $\gamma_k = \frac{\rho\|\mathbf{h}_k\|^2\cos^2\theta_k}{1+\rho\|\mathbf{h}_k\|^2\sin^2\theta_k}$:

Lemma 1 (Distribution of interference): Under the quantization cell approximation model (13), the interference term $\|\mathbf{h}_k\|^2 \sin^2\theta_k$ has a Gamma distribution[5] with parameters $(M-1, \delta)$.

Proof: See Appendix I. ■

Lemma 2 (Joint distribution of signal and interference): Consider two independent Gamma random variables $X \sim \text{Gamma}(1,1)$ and $Y \sim \text{Gamma}(M-1,1)$, and define $I = \delta Y$ and $S = X + (1-\delta)Y$. Then, under the quantization cell approximation (13), the joint distribution of $(\|\mathbf{h}_k\|^2\sin^2\theta_k, \|\mathbf{h}_k\|^2\cos^2\theta_k)$ is the same as that of (I, S).

Proof: See Appendix II. ■

Therefore, the interference term $\rho\|\mathbf{h}_k\|^2\sin^2\theta_k$ has a Gamma distribution $\text{Gamma}(M-1, \rho\delta)$, and the received signal power $\rho\|\mathbf{h}_k\|^2\cos^2\theta_k$ is described as the sum of two independent Gamma variables $\text{Gamma}(1,\rho) + \text{Gamma}(M-1, \rho(1-\delta))$. Note that the signal and interference powers are correlated through Y.

Lemma 3 (Distribution of SINR): Consider the random variables X, Y, I, and S defined in Lemma 2, and define

$$\gamma = \frac{\rho S}{1+\rho I} = \frac{\rho\,(X + (1-\delta)Y)}{1+\rho\delta Y}. \tag{36}$$

Then, under the distribution (13), γ_k and γ have identical distribution with their CDF for $x \geq \frac{1}{\delta} - 1$ given by[6]

$$F_\gamma(x) = 1 - \frac{2^B e^{-\frac{x}{\rho}}}{(x+1)^{M-1}}, \qquad x \geq \frac{1}{\delta} - 1 = 2^{\frac{B}{M-1}} - 1. \tag{37}$$

Proof: See Appendix III. ■

Lemma 3 gives us the distribution of the SINR of any given user (under the condition of semi-orthogonal user selection). To find the distribution of the SINR of a *selected* user $\pi(i) \in \mathcal{S}$, we note that in the user selection process described in Subsection II-C, the ith selected user, $\pi(i)$, is given as $\pi(i) = \arg\max_{k\in\mathcal{A}_{i-1}} \gamma_k$, i.e., the ith user has the maximum SINR among $|\mathcal{A}_{i-1}|$ i.i.d. users. Therefore, it is necessary for us to characterize the behavior of the maximum of a number of i.i.d. random variables. In particular, we analyze the asymptotic case when the number of users K is large, for which the extreme value theory [19]–[21], [2, Appendix A], [8, Appendix II] is useful.

[4]Note that SINR_k in (28) and the sum-rate bound in (33) are valid only for $|\mathcal{S}| = M$. (It is shown that under the semi-orthogonal user selection, $|\mathcal{S}| = M$ with high probability at large K. See [6], [7], [27] for detailed probability analysis.) However, (33) does not preclude the possibility of achieving the multiuser diversity gain alone. Clearly, when $|\mathcal{S}| = 1$ (i.e. when one gives up the multiplexing gain), the system is limited only by noise and fully benefits from multiuser diversity.

[5]A gamma distributed random variable X with parameters (k, θ), for a positive integer k and a positive real number θ, is given by the sum of k i.i.d. exponential random variables with the same parameter θ. Its PDF and CDF are given by

$$f_X(x) = x^{k-1}\frac{e^{-x/\theta}}{(k-1)!\ \theta^k} \tag{34}$$

$$F_X(x) = \frac{\gamma(k, x/\theta)}{\Gamma(k)} = 1 - e^{-x/\theta}\sum_{i=0}^{k-1}\frac{(x/\theta)^i}{i!}. \tag{35}$$

When $\theta = 1$, X reduces to χ^2_{2k}, a chi-square distributed random variable with $2k$ degrees of freedom. (A usual definition for χ^2_{2k} is with $\theta = 2$, but here we use a slightly different definition, as it is often more convenient to have $\theta = 1$ when dealing with complex Gaussian channels with unit variance.)

[6]$F_\gamma(x)$ for $x < \frac{1}{\delta} - 1$ can also be found, but its expression is more involved.

Theorem 1: The jth largest order statistic among $\gamma_1, \cdots, \gamma_K$, denoted as $\gamma_{j:K}$, satisfies

$$P\left\{|\gamma_{j:K} - b_K| \le \rho \log\log\sqrt{K}\right\} \ge 1 - O\left(\frac{1}{\log K}\right), \quad (38)$$

where

$$b_K = \rho \log \frac{2^B K}{\rho^{M-1}} - \rho(M-1)\log\log\frac{2^B K}{\rho^{M-1}}. \quad (39)$$

Therefore, for large K,

$$\gamma_{j:K} = \rho \log \frac{2^B K}{\rho^{M-1}} + O(\log\log K). \quad (40)$$

Proof: See Appendix IV. ∎

By the law of large numbers, the ratio $|\mathcal{A}_i|/K$ converges to some constant α_i, where $\alpha_0 = 1$ since $|\mathcal{A}_0| = K$, and $\alpha_i \ge I_{\epsilon^2}(i, M-i)$, $1 \le i \le M$ with $I_z(a,b)$ the regularized incomplete beta function [7]. Thus, the sum-rate is given by

$$\begin{aligned}\mathbb{E}\{R\} &\approx \mathbb{E}\left\{\sum_{i=1}^{M} \log_2\left(1 + \gamma_{i:|\mathcal{A}_{i-1}|}\right)\right\} \\ &\approx \sum_{i=1}^{M} \log_2\left(1 + \rho \log \frac{2^B K \alpha_{i-1}}{\rho^{M-1}}\right). \end{aligned} \quad (41)$$

Factoring $\gamma_{i:|\mathcal{A}_{i-1}|}$ into the SNR part ρ and the logarithmic term $\Delta \triangleq \log \frac{2^B K \alpha_{i-1}}{\rho^{M-1}}$, we can interpret the latter as the SNR improvement (degradation) factor which includes both the effect of quantization error and multiuser diversity. From the expression we can observe the following:

1) Multiuser diversity of an SNR improvement by a factor of $\log K$ [2], [4], [5] is still preserved under quantized CDI feedback.
2) The quantities 2^B and K are interchangeable. Thus, for a target sum-rate, every doubling of the number of users saves one feedback bit per user.
3) For a target SNR improvement (degradation), B and K should scale with P such that

$$B + \log_2 K = (M-1)\log_2 P + c, \quad (42)$$

for some constant c. That is, for a fixed K, every doubling (3dB increase) of power requires $M-1$ additional feedback bits. This is consistent with the result in (23) for the case of $K = M$. Alternatively, for a fixed B, K could scale with P as $K \propto P^{M-1}$ for a target Δ, or both B and K could be adjusted simultaneously to meet (42). This scaling result extends (23) to the more general case of $K \ge M$.

It is important to note that the formulas (38)-(42) are valid only when both K and $\frac{2^B K}{\rho^{M-1}}$ are large. Moreover, the CDF (37) is valid only for $\gamma_{i:|\mathcal{A}_{i-1}|} \ge 2^{\frac{B}{M-1}} - 1$. Some of these conditions may fail when either B or P is large to the degree that a given K is not large enough to satisfy the conditions. Henceforth, we say that the system is in a *large user regime* if the parameters (K, B, P) are such that the formulas (38)-(42) are valid. Below we qualitatively describe the condition for which the system is in the large user regime.

- For any finite B and P, the system eventually enters a large user regime as $K \to \infty$.
- For a finite K, if P is too large, then $\frac{2^B K}{\rho^{M-1}}$ becomes small, and the system is not in a large user regime. This *high SNR or interference-limited regime* is discussed in Subsection V-A.
- For a finite K, if either B is too large or P is too small, then the condition $\gamma_{i:|\mathcal{A}_{i-1}|} \ge 2^{\frac{B}{M-1}} - 1$ fails, and the system is not in a large user regime. This *high resolution or noise-limited regime* is discussed in Subsection V-B.

A. High SNR or interference-limited regime

In this regime the SINR becomes

$$\lim_{P\to\infty} \gamma_k = \frac{\cos^2\theta_k}{\sin^2\theta_k} = \tilde{\gamma}_k, \quad (43)$$

whose extremal value is given by the following theorem:

Theorem 2: The jth largest order statistic among $\tilde{\gamma}_1, \cdots, \tilde{\gamma}_K$, denoted as $\tilde{\gamma}_{j:K}$, satisfies

$$\begin{aligned}P\Big\{ &\log(2^B K) - \log\log\sqrt{K} \le (M-1)\log(1+\tilde{\gamma}_{j:K}) \\ &\le \log(2^B K) + \log\log\sqrt{K}\Big\} \ge 1 - O\left(\frac{1}{\log K}\right). \end{aligned} \quad (44)$$

Thus, for a large K,

$$\log(1+\tilde{\gamma}_{j:K}) = \frac{1}{M-1}\log(2^B K) + O(\log\log K). \quad (45)$$

Proof: See Appendix V. ∎

The sum-rate then becomes

$$\begin{aligned}\mathbb{E}\{R\} &\approx \mathbb{E}\left\{\sum_{i=1}^{M} \log_2\left(1 + \tilde{\gamma}_{j:K\alpha_{i-1}}\right)\right\} \\ &\approx \frac{M}{M-1}(B + \log_2 K) + \frac{\sum_{i=1}^{M}\log_2\alpha_{i-1}}{M-1}. \end{aligned} \quad (46)$$

We again observe the interchangeability between 2^B and K. Under finite B and K, however, we see that the sum-rate eventually converges to a constant value (46) as $P \to \infty$. This is because the system is interference limited at high SNR due to the unavoidable effect of quantization error. The limiting sum-rate (46), however, grows linearly (ignoring the additive term) with $B + \log_2 K$. In particular, the multiuser diversity amounts to a logarithmic increase to the sum-rate. This is in contrast to previous findings that the sum-rate increase by the multiuser diversity is only by a factor of $\log\log K$ [3]. Therefore, multiuser diversity is even more beneficial in this regime.

B. High resolution or noise-limited regime

As the quantization resolution B goes to infinity, $\theta_k \to 0$, and (28) reduces to $\gamma_k = \rho\|\mathbf{h}_k\|^2$. Also, (28) reduces to $\gamma_k = \rho\|\mathbf{h}_k\|^2$ as $P \to 0$. The random variable $\|\mathbf{h}_k\|^2$ is χ^2_{2M} distributed, and for K i.i.d. χ^2_{2M} random variables, it has been shown that their jth order statistic behaves like $\log K$ for a large K [2], [8]. Thus, $\gamma_{j:K} = \rho\log K + O(\log\log K)$, and

the sum-rate becomes

$$\begin{aligned}\mathbb{E}\{R\} &\approx \mathbb{E}\left\{\sum_{i=1}^{M}\log_2\left(1+\tilde{\gamma}_{j:K\alpha_{i-1}}\right)\right\}\\ &\approx \sum_{i=1}^{M}\log_2\left(1+\rho\log K\alpha_{i-1}\right)\\ &= O\left(M\log\log K\right)\end{aligned} \tag{47}$$

We see that the exchangeability between 2^B and K is no longer observed, i.e. in the high resolution regime, doubling the number of users is worth more than one additional feedback bit.

For finite B and K, we empirically find that

$$(B-B_0)+r\log_2\left(\frac{K}{K_0}\right)=(M-1)\log_2\left(\frac{P}{P_0}\right), \tag{48}$$

where $r\in[1,\infty)$ dictates how many feedback bits doubling the number of users is worth. In the large user regime where $r=1$, (48) reduces to (42). In the limit of high resolution regime ($B\to\infty$ or $P\to 0$), $r\to\infty$. More discussion on this will follow with numerical examples in Section VI.

C. Geometric interpretation

In this subsection we give a geometric intuition for the exchangeability between 2^B and K. Consider the QUB depicted in Figure 2. Suppose user k's normalized channel $\tilde{\mathbf{h}}_k$ is quantized to $\mathbf{c}_i$. Conditioned on this, since the channel is isotropic, the endpoint of the vector $\tilde{\mathbf{h}}_k$ is uniform over the spherical cap of $\mathcal{R}_i$.[7] Now, consider K identical copies of Figure 2, each associated with a user $k\in\{1,\cdots,K\}$. In the interference-limited regime, the SINR $\tilde{\gamma}_k$ only depends on the angle θ_k. Hence, in the user selection process, the user with the smallest θ_k will be chosen. Denote the chosen user as $k^\star$ and its quantization cell $\mathcal{R}_i^\star$. Then, on average, the endpoint of $\tilde{\mathbf{h}}_{k^\star}$ is expected to lie either inside or around the boundary of a smaller spherical cap of an area $2^{-B}/K=2^{-(B+\log_2 K)}$. This new spherical cap corresponds to a quantization cell of a larger codebook of size $2^{B+\log_2 K}=2^BK$. Therefore, either changing the number of users or changing the codebook size has the same effect on the final performance.

The relation between 2^B and K is seen more easily in the RVQ. Since each user has 2^B random codewords and there are K users, effectively there are 2^BK random codewords to choose from in the system. Thus, what matters in the end is the size 2^BK of the *system-wide codebook*.

The above interpretation is for the interference-limited regime. In the large user regime, the chosen user is not necessarily the one with the smallest θ_k. However, somewhat surprisingly, the geometric interpretation seems to continue to hold for the large user regime.

D. Relation to random beamforming (RBF)

In the random beamforming (RBF) scheme proposed in [2], M orthogonal random beams are generated at the transmitter. Then, each user calculates its SINR for each of the M beams and feeds back the maximum SINR value (without quantization) along with a corresponding beam index, after which the transmitter chooses the best user for each beam. Note that $\lceil\log_2 M\rceil$ bits are required for feeding back a user's beam index. Now, consider our limited feedback system that employs a randomly generated optimal codebook of size $N=M$ for CDI quantization. Since the optimal codebook design for $N\le M$ is a set of orthonormal vectors [12], [28], this codebook is equivalent to the random beamformer. Also, note that both systems assume perfect SINR feedback. Therefore, the RBF scheme is essentially equivalent to our limited feedback scheme with $N=M$ and $\phi=0$ with a fixed random orthonormal codebook (without the random rotation by $\mathbf{U}_k$). The similarity can also be observed by comparing $F_\gamma(x)$ in (37) with $F_s(x)$ in [2, eq (15)]. Thus, our scheme can be understood as a generalization of the RBF to the case of $N>M$ and to the beamformers which are not necessarily orthonormal.

[7] In fact, the distribution (13) is derived from the uniformity of $\tilde{\mathbf{h}}_k$ over $\mathcal{R}_i$.

E. CQI quantization

The analysis so far has assumed unquantized feedback of CQI. In practical systems, CQI would have to be quantized appropriately, with the quantization level and thresholds chosen according to the feedback load and the number of supported transmission modes. Since this problem could be rather heuristic, it is not further pursued in this paper. However, we note that when K is asymptotically large, a very simple $(M+1)$-level quantization of γ_k into $[0,g_{M-1}]$, $[g_{M-1},g_{M-2}]$, $\cdots$, $[g_1,g_0]$, $[g_0,\infty)$, where g_j is based on (38)-(39) as $g_j=b_{K\alpha_j}-\rho\log\log\sqrt{K\alpha_j}$, achieves (41) with only $\lceil\log_2(M+1)\rceil$ feedback bits [8], [27]. In practical scenarios, however, the value of K will be much smaller, and an appropriate quantization scheme needs to be developed considering the tradeoff in bit allocations between CQI and CDI quantization. In general, we expect the number of bits for quantizing CQI can be kept relatively small, compared to that required for CDI quantization.

VI. Numerical Results

In this section we present numerical results. We use $M=4$ base-station antennas and $\epsilon=0.25$ for semi-orthogonality threshold throughout this section.

In Figure 4 we compare the sum-rate performances of our proposed scheme, under the quantization upper-bound (QUB) in (13) and the random vector quantization (RVQ), for $K=100$ users and $B=12$ CDI quantization bits per user. For CQI feedback, we use the SINR feedback model of Section V. As was mentioned in Subsection II-E, these two codebook models give performance upper and lower bounds, respectively. For comparison we also show the sum capacity of dirty-paper coding (DPC), and the sum-rate under perfect CSIT. As can be seen from the figure, the sum-rate difference between the two quantization codebook models is very small. Hence, the analysis based on the QUB in the previous sections as well as numerical values from either of the performance curves can serve as an accurate indicator of the sum-rate performance of a general, well-designed codebook. Thus, in

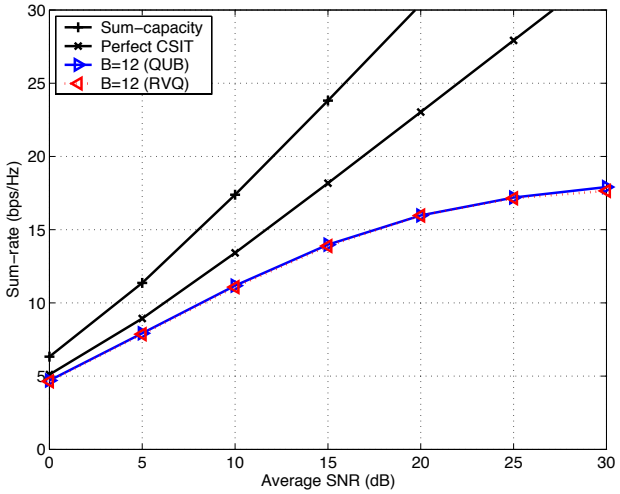

Fig. 4. Sum-rate R versus average SNR P under $M = 4$, $K = 100$, and $B = 12$, for the quantization upper-bound (QUB) and the random vector quantization (RVQ). In both cases SINR-based CQI is used.

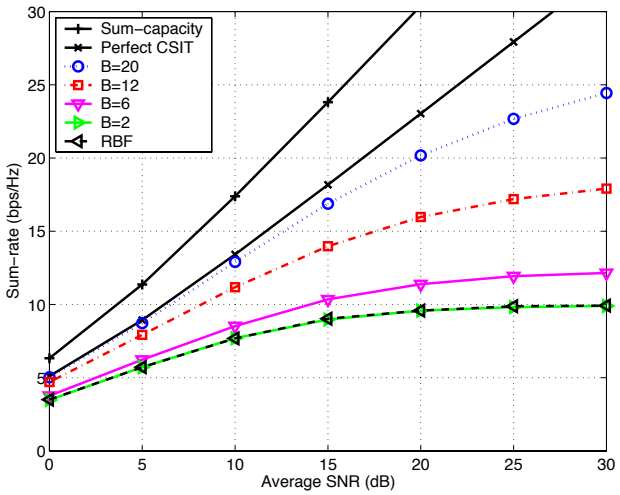

Fig. 5. Sum-rate R versus average SNR P under $M = 4$, $K = 100$, and various B. SINR-based CQI is used.

the remainder of the numerical results, we only plot the sum-rate performance for the QUB for the sake of simplicity of presentation. In all cases, however, we have verified that the two codebook models always give tight performances.

In Figure 5 we present the sum-rate R vs. average SNR P for the system with $K = 100$ users, SINR feedback, and various CDI quantization levels $B = 2$, 6, 12, and 20 bits. We also plot the sum-capacity of DPC, the sum-rate under perfect CSIT, and the sum-rate of the random beamforming (RBF). For $B > \log_2 M = 2$, we use the QUB. For $B = 2$, we use orthonormal codewords, which is optimal [12], [28]. From the figure it is seen that the sum-rate approaches the perfect CSIT sum-rate as B increases. Also we observe that the sum-rate with $B = \log_2 M = 2$ is the same as that of the RBF, which confirms the discussion in Subsection V-D. More importantly, notice that the performance of the RBF (or with $B = 2$) is quite poor compared to the sum-rate under perfect CSIT. Our limited feedback scheme provides a way to use more CDI feedback bits to fill the gap between the performance of the RBF and that under perfect CSIT.

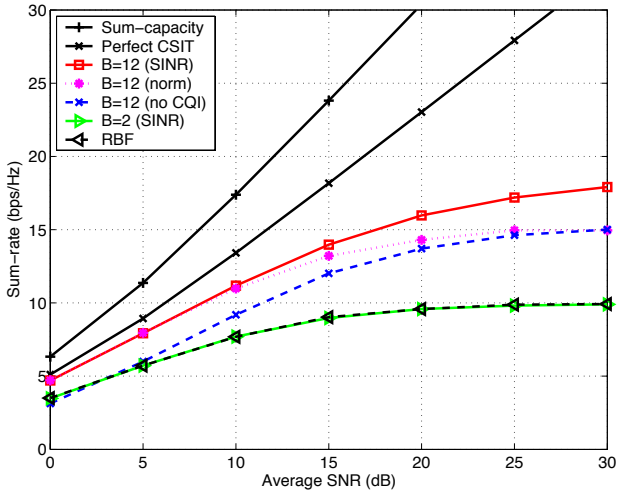

Fig. 6. Sum-rate R versus average SNR P under $M = 4$, $K = 100$, $B = 12$, and various CQI feedback types.

In Figure 6 we compare the sum-rate performances under three different feedback schemes: (A) SINR ($\gamma_k(\phi)$) feedback, (B) channel norm ($\|\mathbf{h}_k\|^2$) feedback, and (C) no CQI feedback. We observe that among the three CQI feedback schemes the SINR feedback performs the best. The channel norm feedback performs close to the SINR feedback at low P (where the system is noise-limited), while at high P (where the system is interference limited) it is only slightly better than having no CQI feedback. This confirms our analysis that the CQI should always be based on SINR, rather than the channel norm. The limiting sum-rate at high SNR for the channel norm feedback is seen to approach 15.2bps/Hz, which is lower than 17.9bps/Hz that is predicted by the upper bound (33). The discrepancy is due to the approximation used in $\gamma(\phi) \approx \gamma(0)$ and the possibility that $|\mathcal{S}| < M$ when K is finite.

Looking at Figure 6, we note that the sum-rate improvement of SINR feedback over no CQI feedback seems to be rather constant over a wide SNR range, whereas in Figure 5 it is seen that the benefit of increasing B becomes more pronounced at high SNR. This implies that a limited feedback resource should be spent more on CQI quantization at low SNR and on CDI quantization at high SNR. As $P \to \infty$ all the sum-rate curves with quantized feedback eventually converge to a finite ceiling given by (46).

In Figure 7 we plot the sum-rate vs. K at $P = 10$ dB. We see that with the SINR feedback the sum-rate benefits from multiuser diversity. With the norm feedback, however, the sum-rate increase is slowed down as K increases and is eventually upper bounded by (33), although the sum-rate increase is maintained longer for a larger B. For $B = 20$ both CQI feedback schemes perform reasonably close to the perfect CSIT case up to 10^4 users.

In Figure 8 we plot the sum-rate vs. B at $P = 10$ dB, for the systems with $K = 100$ users and $K = 400$ users, respectively. It is observed that at low quantization resolution ($B < 6$), the system with $K = 400$ users requires 2 less CDI

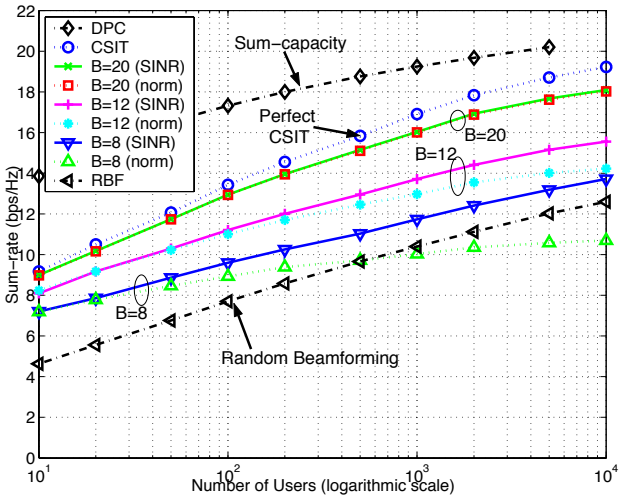

Fig. 7. Sum-rate R versus the number of users K under $M = 4$, $P = 10$, and various B and CQI types.

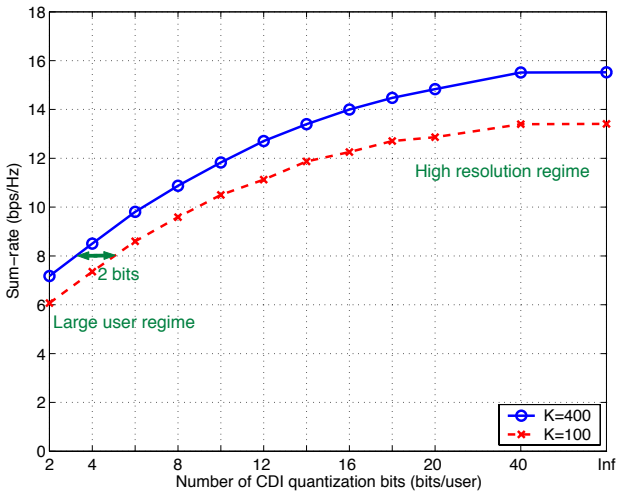

Fig. 8. Sum-rate R versus the number of CDI feedback bits per user B under $M = 4$, $P = 10$ dB, and SINR-based CQI feedback. Note that the scale along the x-axis is not uniform.

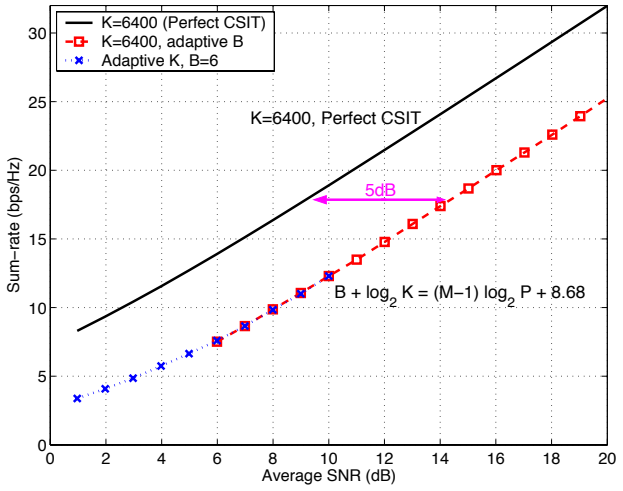

Fig. 9. Sum-rate R versus average SNR P under $M = 4$ and adaptive B and K such that $B + \log_2 K = (M-1)\log_2 P + 8.68$.

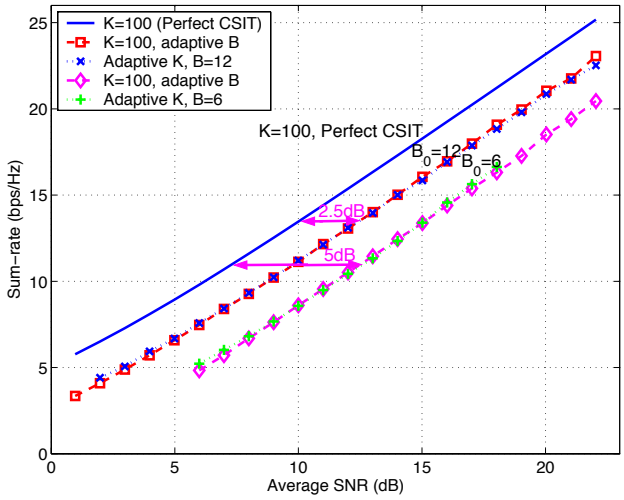

Fig. 10. Sum-rate R versus average SNR P under $M = 4$ and adaptive B and K in a high resolution regime.

feedback bits for the same sum-rate compared to the system with $K = 100$ users, which agrees well with the result in (42) that each doubling of the number of users translates to one additional CDI feedback bit. As B increases, however, this relation ceases to hold, since the system enters a high-resolution regime, where doubling the number of users is worth more than one additional feedback bit, as was discussed in Subsection V-B. With $B = 20$, the sum-rate almost reaches perfect-CSIT sum-rate ($B = \infty$).

To further investigate the interchangeability between 2^B and K, in Figures 9 and 10 we adapt B and K as

$$(B - B_0) + r \log_2 \left(\frac{K}{K_0}\right) = (M-1)\log_2\left(\frac{P}{P_0}\right), \quad (49)$$

so that a constant gap from the perfect CSIT sum-rate is maintained. In Figure 9 we consider a system in a large user regime ($K_0 = 6400$, $B_0 = 6$, $P_0 = 10$). With $r = 1$, this corresponds to

$$B + \log_2 K = (M-1)\log_2 P + 8.68. \quad (50)$$

By either fixing $K = K_0$ and adapting B as a function of P, or fixing $B = B_0$ and adapting K as a function of P, we achieve an SNR gap of about 5dB from the perfect CSIT curve, confirming the third observation in Section V. As we move to higher B or smaller K, the system enters a high resolution regime. To demonstrate this, in Figure 10, we use $K_0 = 100$, $B_0 = 6$, and $P_0 = 10$, for which $r = 1.3$ gives a constant gap of 5dB from the perfect CSIT. When the resolution is even higher ($K_0 = 100$, $B_0 = 12$, and $P_0 = 10$), we need a higher $r = 2.5$ and achieve a smaller gap of 2.5dB. Thus, as is noted in Subsection V-B, in a high resolution regime doubling the number of users worth $r > 1$ feedback bits.

VII. Conclusion

We have investigated a multiuser multi-antenna downlink system under partial channel knowledge at the transmitter, when there are more users than transmit antennas. The SINR distributions and the sum-rates under quantized channel direction information (CDI) and various channel quality informa-

tion (CQI) are derived. We have shown that CDI alone does not achieve the full multiplexing and multiuser diversity gain simultaneously. Our results indicate that to achieve both gains CQI feedback is necessary, and that CQI should be the SINR rather than just the channel magnitude, since SINR captures both the channel magnitude and the quantization error. This implies that any quantization should be applied to SINR rather than directly to the channel magnitude. We have derived tradeoffs between the number of feedback bits, the number of users, and SNR. In particular, for a target performance, having more users reduces feedback load. The number of users and feedback bits can be simultaneously adapted to achieve a constant gap from the perfect CSIT sum-rate.

Future research directions may include investigating CQI quantization and bit allocations between CQI and CDI. Joint quantization of channel directions and SINR may improve performance. Fairness consideration among heterogeneous users, and quantization and feedback strategies for multi-antenna users are also important directions to extend this work.

APPENDIX I
PROOF OF LEMMA 1

The CDF of $\|\mathbf{h}_k\|^2 \sin^2\theta_k$ may be expanded as

$$P\left(\|\mathbf{h}_k\|^2 \sin^2\theta_k \le x\right) = \int_0^\infty F_{\sin^2\theta_k}\left(\frac{x}{y}\right) f_{\|\mathbf{h}_k\|^2}(y)\, dy \tag{51}$$

Using (13), the above becomes

$$\begin{aligned} &P\left(\|\mathbf{h}_k\|^2 \sin^2\theta_k \le x\right) \\ &= \int_0^{\frac{x}{\delta}} f_{\|\mathbf{h}_k\|^2}(y)\, dy + \int_{\frac{x}{\delta}}^\infty 2^B \left(\frac{x}{y}\right)^{M-1} f_{\|\mathbf{h}_k\|^2}(y)\, dy \\ &= F_{\|\mathbf{h}_k\|^2}\left(\frac{x}{\delta}\right) + \left(\frac{x}{\delta}\right)^{M-1} \int_{\frac{x}{\delta}}^\infty \frac{1}{y^{M-1}} f_{\|\mathbf{h}_k\|^2}(y)\, dy. \end{aligned} \tag{52}$$

Noting that $\|\mathbf{h}_k\|^2$ has a Gamma$(M,1)$ distribution, the above can be written as

$$\begin{aligned} &P\left(\|\mathbf{h}_k\|^2 \sin^2\theta_k \le x\right) \\ &= 1 - e^{-x/\delta} \sum_{i=0}^{M-1} \frac{(x/\delta)^i}{i!} + \left(\frac{x}{\delta}\right)^{M-1} \int_{\frac{x}{\delta}}^\infty \frac{e^{-y}}{(M-1)!}\, dy \\ &= 1 - e^{-x/\delta} \sum_{i=0}^{M-2} \frac{(x/\delta)^i}{i!}, \end{aligned} \tag{53}$$

which is the CDF of Gamma$(M-1,\delta)$.

APPENDIX II
PROOF OF LEMMA 2

First, we derive the joint distribution of $I_k = \|\mathbf{h}_k\|^2 \sin^2\theta_k$ and $S_k = \|\mathbf{h}_k\|^2 \cos^2\theta_k$. The joint distribution of (I_k, S_k) is related to that of $(\|\mathbf{h}_k\|^2, \sin^2\theta_k)$ by the transformation

$$f_{\|\mathbf{h}_k\|^2,\sin^2\theta_k}(r,w) = |J|\ f_{I_k,S_k}(u,v), \tag{54}$$

where $u = rw$, $v = r(1-w)$, and the Jacobian of the transformation is given by

$$J = \det \begin{bmatrix} \partial u/\partial r & \partial u/\partial w \\ \partial v/\partial r & \partial v/\partial w \end{bmatrix} = -r. \tag{55}$$

For the channel $\mathbf{h}_k$ with zero-mean i.i.d complex Gaussian entries, its magnitude $\|\mathbf{h}_k\|$ and direction $\tilde{\mathbf{h}}_k = \mathbf{h}_k/\|\mathbf{h}_k\|$ are independent. Since the quantization error θ_k is determined by only the channel direction $\tilde{\mathbf{h}}_k$, it follows that $\|\mathbf{h}_k\|^2$ and $\sin^2\theta_k$ are independent. Using the fact that $\|\mathbf{h}_k\|^2$ follows Gamma$(M,1)$ distribution, and that the distribution of $\sin^2\theta_k$ is given by (13), we can write the joint distribution of (I_k, S_k) as

$$\begin{aligned} f_{I_k,S_k}(u,v) &= \frac{1}{|J|} f_{\|\mathbf{h}_k\|^2,\sin^2\theta_k}(r,w) \\ &= \frac{1}{r}\ f_{\|\mathbf{h}_k\|^2}(r)\ f_{\sin^2\theta_k}(w) \\ &= \begin{cases} \frac{1}{r}\ \frac{e^{-r} r^{M-1}}{(M-1)!}\ 2^B (M-1) w^{M-2}, & 0 \le w \le \delta \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} 2^B\ \frac{u^{M-2} e^{-u}}{(M-2)!}\ e^{-v}, & 0 \le \frac{u}{u+v} \le \delta \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{56}$$

Now, consider two independent Gamma random variables $X \sim$ Gamma$(1,1)$ and $Y \sim$ Gamma$(M-1,1)$, and define $I = \delta Y$ and $S = X + (1-\delta)Y$. We would like to show that the joint distribution $f_{I,S}(u,v)$ is also given by (56). Using the linear relation between (X,Y) and (I,S), we can write

$$f_{I,S}(u,v) = \frac{1}{\delta}\ f_X\left(\left(1-\frac{1}{\delta}\right)u + v\right) f_Y\left(\frac{1}{\delta}u\right). \tag{57}$$

For (u,v) such that $\left(1-\frac{1}{\delta}\right)u + v \ge 0$, the above becomes

$$\begin{aligned} f_{I,S}(u,v) &= \frac{1}{\delta} e^{-\left((1-\frac{1}{\delta})u+v\right)} \frac{\left(\frac{u}{\delta}\right)^{M-2} e^{-\frac{u}{\delta}}}{(M-2)!} \\ &= 2^B\ \frac{u^{M-2} e^{-u}}{(M-2)!}\ e^{-v}. \end{aligned} \tag{58}$$

For (u,v) such that $\left(1-\frac{1}{\delta}\right)u + v < 0$, clearly $f_{I,S}(u,v) = 0$, since X can not take negative values. Therefore, the joint distribution $f_{I,S}(u,v)$ is the same as (56).

APPENDIX III
PROOF OF LEMMA 3

It is clear from Lemma 2 that $\gamma_k = \frac{\rho S_k}{1+\rho I_k}$ and $\gamma = \frac{\rho S}{1+\rho I}$ have identical distribution. Therefore, it suffices to derive the CDF of γ. The probability $P(\gamma \ge x)$ is given by

$$\begin{aligned} P(\gamma \ge x) &= P\left(\frac{\rho X + \rho(1-\delta)Y}{1+\rho\delta Y} \ge x\right) \\ &= \int_0^\infty P\left(X \ge \frac{x}{\rho} + (\delta x + \delta - 1)y\right) f_Y(y)\, dy. \end{aligned} \tag{59}$$

If $x \ge \frac{1-\delta}{\delta}$, then $\frac{x}{\rho} + (\delta x + \delta - 1)y$ is nonnegative for any $y \ge 0$, and noting that $X \sim$ Gamma$(1,1)$ and $Y \sim$ Gamma$(M-1,1)$, the above can be written as

$$\begin{aligned} P(\gamma \ge x) &= \int_0^\infty e^{-\frac{x}{\rho} - (\delta x + \delta - 1)y} \frac{y^{M-2} e^{-y}}{(M-2)!}\, dy \\ &= \frac{e^{-\frac{x}{\rho}}}{\delta^{M-1}(x+1)^{M-1}} \underbrace{\int_0^\infty \frac{y^{M-2} e^{-(\delta x+\delta)y}}{(M-2)!\left(\frac{1}{\delta x+\delta}\right)^{M-1}}\, dy}_{=1} \\ &= \frac{2^B e^{-\frac{x}{\rho}}}{(x+1)^{M-1}}, \quad x \ge \frac{1-\delta}{\delta}. \end{aligned} \tag{60}$$

where in evaluating the integral we used the observation that its integrand is the PDF of Gamma$(M-1, \frac{1}{\delta x+\delta})$. From this, the CDF (37) immediately follows.

APPENDIX IV
PROOF OF THEOREM 1

For a complete proof, we apply Theorem 3 in the Appendix VI to the distribution (13) with $a_K = \rho$ and b_K as in (39):

$$\begin{aligned} K(1-F_\gamma(a_K z+b_K)) &= \frac{K2^B \exp\left(-\frac{a_K z+b_K}{\rho}\right)}{(a_K z+b_K+1)^{M-1}} \\ &= \frac{K2^B \exp\left(-z-\log\frac{2^B K}{\rho^{M-1}}+(M-1)\log\log\frac{2^B K}{\rho^{M-1}}\right)}{\left(\rho z+\rho\log\frac{2^B K}{\rho^{M-1}}-\rho(M-1)\log\log\frac{2^B K}{\rho^{M-1}}+1\right)^{M-1}} \\ &= \frac{e^{-z}\left(\rho\log\frac{2^B K}{\rho^{M-1}}\right)^{M-1}}{\left(\rho z+\rho\log\frac{2^B K}{\rho^{M-1}}-\rho(M-1)\log\log\frac{2^B K}{\rho^{M-1}}+1\right)^{M-1}} \\ &\to e^{-z} \quad \text{as} \quad K\to\infty. \end{aligned} \tag{61}$$

Therefore, $F_\gamma(x)$ belongs to the type (iii) distribution of the Appendix VI. Then, Theorem 1 can be proved by utilizing Theorem 4 in the Appendix VI, in a similar manner used in [8] to prove Lemma 4 of the Appendix VI.

Alternatively, the Theorem can be directly proved applying Lemma 4 with the change of variable $\gamma' = \frac{\gamma+1}{\rho}$. Then,

$$\begin{aligned} F_{\gamma'}(z) &= F_\gamma(\rho z-1) \\ &= 1-\frac{2^B e^{1/\rho}}{\rho^{M-1}} z^{-(M-1)} e^{-z}, \quad z\geq\frac{1}{\delta\rho}. \end{aligned} \tag{62}$$

Therefore, Lemma 4 applies with

$$\alpha = M-1, \quad \frac{1}{\beta} = \frac{2^B e^{1/\rho}}{\rho^{M-1}}. \tag{63}$$

From this it is straightforward to obtain (38)-(39). It is instructive to note that the choice of b_K in (39) may be inferred from solving $1-F_\gamma(x) = \frac{1}{K}$ for x.

APPENDIX V
PROOF OF THEOREM 2

We use Theorem 3 in Appendix VI. Let $a_K = (2^B K)^{\frac{1}{M-1}}$ and $b_K = -1$, and consider the distribution (30). When $z>0$, $a_K z+b_K \geq \frac{1}{\delta}-1$ for a sufficiently large K, and we have

$$\begin{aligned} K\left(1-F_{\tilde{\gamma}_k}(a_K z+b_K)\right) &= \frac{K2^B}{(a_K z+b_K+1)^{M-1}} \\ &= z^{-(M-1)}, \quad z>0 \end{aligned} \tag{64}$$

When $z\leq 0$, $a_K z+b_K<0$ and

$$K\left(1-F_{\tilde{\gamma}_k}(a_K z+b_K)\right) = 0. \tag{65}$$

Therefore, $F_{\tilde{\gamma}_k}$ satisfies (83) with $l=1$ and $\epsilon = M-1$, i.e. $F_{\tilde{\gamma}_k}$ belongs to the type (i) distribution of the Appendix VI.

We now apply Theorem 4 in Appendix VI. First, we evaluate functions defined in Theorem 4 as

$$\delta_K(z) = 1-F_{\tilde{\gamma}_k}(a_K z+b_K) = \frac{1}{K z^{M-1}}, \tag{66}$$

and

$$g(j, K\delta_K(z)) = \begin{cases} e^{-\frac{1}{z^{(M-1)}}}, & j=1 \\ e^{-\frac{1}{z^{(M-1)}}}\left(\frac{1}{(j-1)!z^{(j-1)(M-1)}}-\frac{1}{(j-2)!z^{(j-2)(M-1)}}\right), & j\geq 2, \end{cases} \tag{67}$$

and for $l=1$,

$$\begin{aligned} \Theta(z) &= \left|\frac{1}{(j-1)!}\int_{K\delta_K(z)}^{-\log\Lambda_1(z)} \varpi^{j-1}e^{-\varpi}\,d\varpi\right| \\ &= \left|\frac{1}{(j-1)!}\int_{z^{-(M+1)}}^{z^{-(M+1)}} \varpi^{j-1}e^{-\varpi}\,d\varpi\right| = 0 \end{aligned} \tag{68}$$

Now, let us evaluate (84) for two different values of z at $z_1 = (\log\sqrt{K})^{1/(M-1)}$ and $z_2 = (\log\sqrt{K})^{-1/(M-1)}$. For $z = z_1 = (\log\sqrt{K})^{1/(M-1)}$, we have $\delta_K(z_1) = \frac{1}{K\log\sqrt{K}}$, and therefore,

$$\begin{aligned} K\delta_K^2(z_1)g(j, K\delta_K(z_1)) &= \frac{1}{K(\log\sqrt{K})^2} g(j, K\delta_K(z_1)) \\ &= o\left(1/K\right) \end{aligned} \tag{69}$$

It can be easily verified that the right-hand side (RHS) of (84) for $z=z_1$ is $O\left(\frac{1}{K^2}\right)$. Therefore, (84) is simplified as

$$\begin{aligned} &\left|F_{j:K}\left(a_K z_1+b_K\right)-\Lambda_1(z_1)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_1))^i}{i!}+o\left(\frac{1}{K}\right)\right| \\ &= O\left(1/K^2\right). \end{aligned} \tag{70}$$

Similarly, at $z=z_2=(\log\sqrt{K})^{-1/(M-1)}$, we have $\delta_K(z_2) = \frac{\log\sqrt{K}}{K}$, and therefore,

$$\begin{aligned} K\delta_K^2(z_2)g(j, K\delta_K(z_2)) &= \frac{(\log\sqrt{K})^2}{K} g(j, K\delta_K(z_2)) \\ &= o\left(1/K\right). \end{aligned} \tag{71}$$

We can easily show that the RHS of (84) for $z=z_2$ is $O\left(\frac{(\log\sqrt{K})^3}{K}\right)$. Then, (84) is simplified as

$$\begin{aligned} &\left|F_{j:K}\left(a_K z_2+b_K\right)-\Lambda_1(z_2)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_2))^i}{i!}+o\left(\frac{1}{K}\right)\right| \\ &= O\left((\log\sqrt{K})^3/K\right). \end{aligned} \tag{72}$$

Furthermore, from

$$\begin{aligned} \Lambda_1(z_1)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_1))^i}{i!} &= e^{-\frac{1}{\log\sqrt{K}}}\sum_{i=0}^{j-1}\frac{1}{i!(\log\sqrt{K})^i} \\ &= 1-O\left(1/\log K\right), \end{aligned} \tag{73}$$

and

$$\begin{aligned} \Lambda_1(z_2)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_2))^i}{i!} &= e^{-\log\sqrt{K}}\sum_{i=0}^{j-1}\frac{\left(\log\sqrt{K}\right)^i}{i!} \\ &= O\left(\left(\log\sqrt{K}\right)^j/\sqrt{K}\right), \end{aligned} \tag{74}$$

we obtain

$$\left|\Lambda_1(z_1)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_1))^i}{i!}-\Lambda_1(z_2)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_1(z_2))^i}{i!}\right| \geq 1-O\left(1/\log K\right). \quad (75)$$

Combining (70), (72), and (75), we obtain

$$\left|F_{j:K}\left(a_K z_1+b_K\right)-F_{j:K}\left(a_K z_2+b_K\right)\right| \geq 1-O\left(\frac{1}{\log K}\right), \quad (76)$$

Plugging in $a_K=(2^B K)^{\frac{1}{M-1}}$ and $b_K=-1$, we have

$$\left|F_{j:K}\left(\left(2^B K\log\sqrt{K}\right)^{\frac{1}{M-1}}-1\right) -F_{j:K}\left(\left(\frac{2^B K}{\log\sqrt{K}}\right)^{\frac{1}{M-1}}-1\right)\right| \geq 1-O\left(1/\log K\right), \quad (77)$$

or equivalently,

$$P\left(\left(\frac{2^B K}{\log\sqrt{K}}\right)^{\frac{1}{M-1}}\leq 1+\tilde{\gamma}_{j:K}\leq\left(2^B K\log\sqrt{K}\right)^{\frac{1}{M-1}}\right) \geq 1-O\left(1/\log K\right), \quad (78)$$

from which (44) readily follows.

APPENDIX VI
USEFUL RESULTS FROM EXTREME ORDER STATISTICS

We review some useful results from extreme order statistics, i.e. the statistics of extreme values of i.i.d. random variables. Let $X_1,\cdots,X_K$ be a sequence of i.i.d random variables with a distribution F_X. Denote as $X_{j:K}$ the jth largest among $X_1,\cdots,X_K$. In this appendix we are interested in the distribution of $X_{j:K}$.

The following Theorem, taken from [20], [8, Theorem 5-6], gives the limiting distribution of $F_{j:K}(x)=P(X_{j:K}\leq x)$:

Theorem 3: Let j be a fixed index, and $a_K>0$ and b_K be sequences of real numbers such that $\lim_{K\to\infty}F_{j:K}(a_K z+b_K)$ exists. Then, the limit has the following form:

$$\lim_{K\to\infty}F_{j:K}(a_K z+b_K)=\Lambda_l(z)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_l(z))^i}{i!}, \quad (79)$$

where the function $\Lambda(z)$ must be one of the following three types:

$$\text{Type }(i)\quad \Lambda_1(z)=\begin{cases}0, & z\leq 0\\ e^{-z^{-\epsilon}}, & z>0,\epsilon>0\end{cases} \quad (80)$$

$$\text{Type }(ii)\quad \Lambda_2(z)=\begin{cases}e^{-(-z)^{\epsilon}}, & z\leq 0,\epsilon>0\\ 1, & z>0\end{cases} \quad (81)$$

$$\text{Type }(iii)\quad \Lambda_3(z)=e^{-e^{-z}}. \quad (82)$$

A necessary and sufficient condition for $F_{j:K}$ to be of type $l=1,2,3$ is

$$\lim_{K\to\infty}K\left(1-F_X(a_K z+b_K)\right)=-\log\Lambda_l(z). \quad (83)$$

The following theorem, taken from [21], [8, Theorem 7], gives the rate of convergence to the limiting distributions:

Theorem 4: Consider F_X, a_K, and b_K that satisfy (79) with type l. If $\frac{1}{2}<F_X(a_K z+b_K)<1$ and $-\log\Lambda_l(z)<\infty$, then for a fixed index j,

$$\left|F_{j:K}(a_K z+b_K)-\Lambda_l(z)\sum_{i=0}^{j-1}\frac{(-\log\Lambda_l(z))^i}{i!} +\frac{1}{2}K\delta_K^2(z)g(j,K\delta_K(z))\right| \leq\frac{\pi}{2}e^{2K\delta_K(z)}K\delta_K^3(z)\left[\frac{4}{3\left(1-2\delta_K(z)\right)}+\left\{\frac{16K\delta_K^3(z)}{9\left(1-2\delta_K(z)\right)^2} +\frac{8K\delta_K^2(z)}{3\left(1-2\delta_K(z)\right)}\right\}\exp\left(K\delta_K^2(z)\left(1+\frac{4\delta_K(z)}{3\left(1-2\delta_K(z)\right)}\right)\right)\right] +\Theta(z), \quad (84)$$

where

$$\delta_K(z)=1-F_X(a_K z+b_K), \quad (85)$$

and

$$g(z,\vartheta)=\begin{cases}0, & z\leq 0\\ e^{-\vartheta}, & 0<z\leq 1\\ e^{-\vartheta}\left(\frac{\vartheta^{\lceil z\rceil-1}}{(\lceil z\rceil-1)!}-\frac{\vartheta^{\lceil z\rceil-2}}{(\lceil z\rceil-2)!}\right), & z>1\end{cases} \quad (86)$$

and

$$\Theta(z)=\left|\frac{1}{(j-1)!}\int_{K\delta_K(z)}^{-\log\Lambda_l(z)}\varpi^{j-1}e^{-\varpi}\,d\varpi\right|. \quad (87)$$

The following lemma, derived in [8, Lemma 6], states the limiting value of $X_{j:K}$ for a particular distribution of our interest:

Lemma 4: The distribution

$$F_X(z)=1-\frac{1}{\beta}z^{-\alpha}e^{-z},\quad \alpha\geq 0,\beta>0 \quad (88)$$

belongs to the type (iii) distribution with

$$a_K=1, \quad (89)$$

$$b_K=\log\frac{K}{\beta}-\alpha\log\log\frac{K}{\beta}, \quad (90)$$

and the jth maximum $X_{j:K}$ satisfies

$$P\left(|X_{j:K}-b_K|\leq\log\log\sqrt{K}\right)\geq 1-O\left(\frac{1}{\log K}\right). \quad (91)$$

This lemma is proved by utilizing theorems 3 and 4.

REFERENCES

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, July 2003.

[2] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, pp. 506–522, Feb. 2005.

[3] ——, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, submitted for publication.

[4] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, June 1995, pp. 331–335.

[5] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.
[6] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 528–541, Mar. 2006.
[7] ——, "Sum-rate optimal multi-antenna downlink beamforming strategy based on clique search," in *Proc. IEEE GLOBECOM*, Nov. 2005, pp. 1510–1514.
[8] M. Maddah-Ali, M. Ansari, and A. Khandani, "An efficient signaling scheme for MIMO broadcast systems: Design and performance evaluation," *IEEE Trans. Inform. Theory*, submitted.
[9] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Commun. Lett.*, vol. 7, pp. 370–372, Aug. 2003.
[10] A. Narula, N. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transimssion over fading," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.
[11] D. J. Love, R. W. H. Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2735–2747, Oct. 2003.
[12] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2562–2579, Oct. 2003.
[13] S. Zhou, Z. Wang, and G. Giannakis, "Quantifying the power loss when transmit beamforming relies on finite rate feedback," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1948–1957, July 2005.
[14] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.
[15] P. Ding, D. J. Jove, and M. D. Zoltowski, "Multiple antenna broadcast channels with partial and limited feedback," *IEEE Trans. Signal Processing*, submitted for publication.
[16] K. Huang, R. W. Heath, and J. G. Andrews, "Joint beamforming and scheduling for SDMA systems with limited feedback," *IEEE Trans. Commun.*, submitted.
[17] H. Viswanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," in *Proc. Allerton Conf. Communications, Control and Computing*, 2001.
[18] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "Finding NEMO: Near mutually orthogonal sets and applications to MIMO broadcast scheduling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2005.
[19] H. David, *Order Statistics*. New York: John Wiley and Sons, Inc., 1980.
[20] N. Smirnov, "Limit distributions for the terms of a variational series," *Trudy Mat. Inst.*, vol. 25, 1949.
[21] W. Dziubdziela, "On convergence rates in the limit laws of extreme order statistics," in *Trans. 7th Prague Conference and 1974 European Meeting of Statisticians*, 1974, pp. 119–127.
[22] N. Jindal, "A feedback reduction technique for MIMO broadcast channels," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, July 2006.
[23] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.
[24] W. Santipach and M. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, July 2004, p. 290.
[25] C. Au-Yeung and D. Love, "On the performance of random vector quantization limited feedback beamforming," preprint.
[26] S. Jafar and S. Srinivasa, "On the optimality of beamforming with quantized feedback," *IEEE Trans. Inform. Theory*, submitted.
[27] C. Swannack, E. Uysal-Biyikoglu, and G. W. Wornell, "MIMO broadcast scheduling with limited channel state information," in *Proc. Allerton Conf. Communications, Control and Computing*, Oct. 2005.
[28] S. Srinivasa and S. Jafar, "Vector channel capacity with quantized feedback," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2005, pp. 2674–2678.

Taesang Yoo (S04) received the B.S. degree (Hons.) in electrical engineering from Seoul National University,Seoul, Korea, in 1998, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2003. He is currently working towards the Ph.D. degree at Stanford University. He was a Summer Intern in the Wireless Communications Group, Lucent Bell Laboratories, Holmdel,NJ, in 2005. In Summer 2002, he was an Intern at Qualcomm, Campbell, CA. From 2000 to 2001, he was with Xeline, Seoul, Korea, where he worked on the design of powerline communications chipsets. He was also an Engineer at DSI, Seoul,Korea, from 1998 to 2000. His research interests include multiple antenna systems, cellular systems, and communication theory.

Nihar Jindal received the B.S. degree in Electrical Engineering and Computer Science from U.C. Berkeley in 1999, and the M.S. and Ph.D. degrees in Electrical Engineering from Stanford University in 2001 and 2004. He is currently an assistant professor at the University of Minnesota. His industry experience includes summer internships at Intel Corporation, Santa Clara, CA in 2000 and at Lucent Bell Labs, Holmdel, NJ in 2002. His research interests include multiple-antenna/multi-user channels, dynamic resource allocation, and sensor and ad-hoc networks. Dr. Jindal was the recipient of the 2005 IEEE Communications Society and Information Theory Society Joint Paper Award.

Andrea Goldsmith is a professor of Electrical Engineering at Stanford University, and was previously an assistant professor of Electrical Engineering at Caltech. She has also held industry positions at Maxim Technologies and at AT&T Bell Laboratories, and is currently on leave from Stanford as cofounder and CTO of Quantenna Communications, Inc. Her research includes work on capacity of wireless channels and networks, wireless communication and information theory, energy-constrained wireless communications, wireless communications for distributed control, and cross-layer design of wireless networks. She is author of the book "Wireless Communications" and co-author of the book "MIMO Wireless Communications," both published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley.

Dr. Goldsmith is a Fellow of the IEEE and of Stanford. She has received several awards for her research, including the National Academy of Engineering Gilbreth Lectureship, the Alfred P. Sloan Fellowship, the Stanford Terman Fellowship, the National Science Foundation CAREER Development Award, and the Office of Naval Research Young Investigator Award. She was also a co-recipient of the 2005 IEEE Communications Society and Information Theory Society joint paper award. She currently serves as associate editor for the IEEE Transactions on Information Theory and as editor for the Journal on Foundations and Trends in Communications and Information Theory and in Networks. She was previously an editor for the IEEE Transactions on Communications and for the IEEE Wireless Communications Magazine, and has served as guest editor for several IEEE journal and magazine special issues. Dr. Goldsmith is active in committees and conference organization for the IEEE Information Theory and Communications Societies and is an elected member of the Board of Governors for both societies. She is a distinguished lecturer for the IEEE Communications Society, the second vice-president and student committee chair of the IEEE Information Theory Society, and the technical program co-chair for the 2007 IEEE International Symposium on Information Theory.