A Feedback Reduction Technique for MIMO Broadcast Channels

Nihar Jindal

Department of Electrical and Computer Engineering

University of Minnesota

Minneapolis, MN 55455, USA

Email: nihar@umn.edu

Abstract— **A multiple antenna broadcast channel with perfect channel state information at the receivers is considered. If each receiver quantizes its channel knowledge to a finite number of bits which are fed back to the transmitter, the large capacity benefits of the downlink channel can be realized. However, the required number of feedback bits per mobile must be scaled with both the number of transmit antennas and the system SNR, and thus can be quite large in even moderately sized systems. It is shown that a small number of antennas can be used at each receiver to improve the quality of the channel estimate provided to the transmitter. As a result, the required feedback rate per mobile can be significantly decreased.**

I. Introduction

In multiple antenna broadcast (downlink) channels, capacity can be tremendously increased by adding antennas at only the access point (transmitter) [1]. However, the transmitter must have accurate channel state information (CSI) in order to realize these multiplexing gains. In frequency-division duplexed systems, training can be used to obtain channel knowledge at each of the mobile devices (receivers), but obtaining CSI at the access point generally requires feedback from each mobile.

In the practically motivated *finite rate feedback* model, each mobile feeds back a finite number of bits regarding its channel instantiation at the beginning of each block or frame. The feedback bits are determined by quantizing the channel vector to one of 2^B quantization vectors. A downlink channel with such a feedback mechanism was analyzed in [2][3][4]. While only a few feedback bits suffice to obtain near-perfect CSIT performance in point-to-point MISO (multiple-input, single-output) channels [5][6], considerably more feedback is required in downlink channels. In fact, the feedback load per mobile must be scaled with the number of transmit antennas as well as the system SNR in downlink channels in order to achieve rates close to those achievable with perfect CSIT. In [2][4], it is shown that the following scaling of feedback bits per mobile

$$B = \frac{M-1}{3} P_{dB} \qquad (1)$$

suffices to maintain a maximum gap of 3 dB between perfect CSIT and limited feedback performance when zero-forcing beamforming is used. This feedback load, however, can be prohibitively large for even reasonable size systems. In a 10

antenna system operating at 10 dB, for example, this equates to 30 feedback bits per mobile.

In this paper, we propose a method that significantly reduces the required feedback load by utilizing a small number of receive antennas (denoted by N) at each mobile. The multiple receive antennas are not used to increase the number of data streams received at each mobile, as they are in point-to-point MIMO systems, but instead are used to improve the quality of the channel estimate provided to the transmitter. Each mobile linearly combines the received signals on its N antennas to produce a scalar output, thereby creating an effective single antenna channel at each mobile. Transmission is then performed as in a multiple transmit antenna, single receive antenna downlink channel. However, the coefficients of the linear combiner at each mobile are not arbitrary, but instead are chosen to produce the effective single antenna channel that can be quantized with minimal error, thereby decreasing the quantization error for each mobile. Increasing the number of receive antennas N increases the space of possible effective channels, and thus leads to reduced quantization error. In a 10 antenna system operating at 10 dB, for example, this method reduces the feedback from 30 bits per mobile in the $N = 1$ scenario to 25 bits and 21 bits, for $N = 2$ and $N = 3$, respectively.

Notation: We use lower-case boldface to denote vectors, upper-case boldface for matrices, and the symbol $(\cdot)^H$ for the conjugate transpose. The norm of vector $\mathbf{x}$ is denoted $||\mathbf{x}||$.

II. System Model

We consider a K receiver multiple antenna broadcast channel in which the transmitter (access point) has M antennas, and each of the receivers has N antennas. The received signal at the i-th antenna is described as:

$$y_i = \mathbf{h}_i^H \mathbf{x} + n_i, \quad i = 1, \ldots, NK \qquad (2)$$

where $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{KN}$ are the channel vectors (with $\mathbf{h}_i \in \mathbb{C}^{M \times 1}$) describing the KN receive antennas, the vector $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmitted signal, and $\mathbf{n}_1, \ldots, \mathbf{n}_{NK}$ are independent complex Gaussian noise terms with unit variance. Note that receiver 1 has access to signals $y_1, \ldots, y_N$, receiver 2 has access to $y_{N+1}, \ldots, y_{2N}$, and the i-th receiver has access to $y_{(i-1)N+1}, \ldots, y_{Ni}$. There is a transmit power constraint of P, i.e. we must satisfy $E[||\mathbf{x}||^2] \leq P$. We use $\mathbf{H}_i$ to

denote the concatenation of the i-th receiver's channels, i.e. $\mathbf{H}_i = [\mathbf{h}_{(i-1)N+1} \cdots \mathbf{h}_{Ni}]$. For simplicity of exposition we assume that the number of mobiles is equal to the number of transmit antennas, i.e., $K = M$. The results can easily be extended to the case where $K < M$, and the proposed technique can be combined with user selection when $K > M$. Furthermore, the number of receive antennas is assumed to be no larger than the number of transmit antennas, i.e., $N \leq M$.

We consider a block fading channel, with independent Rayleigh fading from block to block. Each of the receivers is assumed to have perfect and instantaneous knowledge of its own channel $\mathbf{H}_i$. Notice it is not required for mobiles to know the channel of other mobiles. In this work we study only the *ergodic capacity*, or the average rates achieved over an infinite number of blocks (or channel realizations).

A. Finite Rate Feedback Model

Here we briefly describe the feedback model for a single receive antenna ($N = 1$). At the beginning of each block, each receiver quantizes its channel (with $\mathbf{h}_i$ assumed to be known perfectly at the i-th receiver) to B bits and feeds back the bits perfectly and instantaneously to the access point. Vector quantization is performed using a codebook $\mathcal{C}$ that consists of 2^B M-dimensional unit norm vectors $\mathcal{C} \triangleq \{\mathbf{w}_1, \ldots, \mathbf{w}_{2^B}\}$, where B is the number of feedback bits. Each receiver quantizes its channel vector to the beamforming vector that forms the minimum angle to it, or equivalently that maximizes the inner product [7] [8]. Thus, user i quantizes its channel to $\hat{\mathbf{h}}_i$, chosen according to:

$$\hat{\mathbf{h}}_i = \arg \max_{\mathbf{w}=\mathbf{w}_1,\ldots,\mathbf{w}_{2^B}} |\mathbf{h}_i^H \mathbf{w}| \quad (3)$$

$$= \arg \min_{\mathbf{w}=\mathbf{w}_1,\ldots,\mathbf{w}_{2^B}} \sin^2(\angle(\mathbf{h}_i, \mathbf{w})). \quad (4)$$

and feeds the quantization index back to the transmitter. It is important to notice that only the direction of the channel vector is quantized, and no magnitude information is conveyed to the transmitter. The quantization error can be thought of as either the angle between the channel and its quantization $\angle(\mathbf{h}_i, \hat{\mathbf{h}}_i)$ or the quantity $\sin^2(\angle(\mathbf{h}_i, \hat{\mathbf{h}}_i))$.

In this work we use *random vector quantization* (RVQ), in which each of the 2^B quantization vectors is independently chosen from the isotropic distribution on the M-dimensional unit sphere [5]. To simplify analysis, each receiver is assumed to use a different and independently generated codebook. We analyze performance averaged over all such choices of random codebooks. Random codebooks are used because the optimal vector quantizer for this problem is not known in general and known bounds are rather loose, whereas RVQ is amenable to analysis and also provides performance that is measurably close to the optimal [5].

B. Zero-Forcing Beamforming

After receiving the quantization indices from each of the mobiles, the AP can use zero-forcing beamforming (ZFBF) to transmit data to the M users. Let us again consider the $N = 1$ scenario, where the channels are the vectors $\mathbf{h}_1, \ldots, \mathbf{h}_M$.

Since the transmitter does not have perfect CSI, ZFBF must be performed based on the quantizations instead of the actual channels. When ZFBF is used, the transmitted signal is defined as $\mathbf{x} = \sum_{i=1}^M x_i \mathbf{v}_i$, where each x_i is a scalar (chosen complex Gaussian with power P/M) intended for the i-th receiver, and $\mathbf{v}_i \in \mathcal{C}^M$ is the beamforming vector for the i-th receiver. The beamforming vectors $\mathbf{v}_1, \ldots, \mathbf{v}_M$ are chosen as the normalized rows of the matrix $[\hat{\mathbf{h}}_1 \cdots \hat{\mathbf{h}}_M]^{-1}$, i.e., they satisfy $||\mathbf{v}_i|| = 1$ for all i and $\hat{\mathbf{h}}_i^H \mathbf{v}_j = 0$ for all $j \neq i$. If all multi-user interference is treated as additional noise, the resulting SINR at the i-th receiver is given by:

$$SINR_i = \frac{\frac{P}{M}|\mathbf{h}_i^H \mathbf{v}_i|^2}{1 + \sum_{j \neq i} \frac{P}{M}|\mathbf{h}_i^H \mathbf{v}_j|^2}. \quad (5)$$

Note that the interference terms in the denominator are strictly positive because $\mathbf{h}_i \neq \hat{\mathbf{h}}_i$, i.e., due to the quantization error.

III. EFFECTIVE CHANNEL QUANTIZATION

In this section we describe the proposed method to reduce the quantization error in the transmitter's estimate of the mobile channels. We begin by first describing a simple, antenna-selection method for reducing feedback, which motivates the better performing effective channel method.

A simple method to utilize N receive antennas is to separately quantize each of the N channel vectors and then feed back the index of only the best of the N quantizations. If, for example, antenna 1 had the minimum quantization error, the mobile would only send the quantization index describing antenna 1 and would only utilize the first antenna when receiving. It is straightforward to show that choosing the best of the N channel quantizations, each from a codebook of size 2^B, is equivalent to quantizing a single channel using a codebook of size $N \cdot 2^B$. Thus, if B feedback bits are sent by each mobile, a system with N antennas per mobile will perform identical to a single receive antenna system with $B + \log_2 N$ feedback bits per mobile. Thus, utilizing N receive antennas in this simple manner decreases the feedback load by $\log_2 N$ bits per mobile.

A more significant decrease in feedback load can be achieved by considering all possible *linear combinations* of the N received signals, instead of limiting the system to selection of one of the N signals. Consider the effective received signal at the first receiver after linearly combining the N received signals by complex weights $\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \ldots, \gamma_{1,N})$ satisfying $|\boldsymbol{\gamma}_1| = 1$:

$$y_1^{\text{eff}} = \sum_{k=1}^N \gamma_{1,k}^H y_k = \sum_{k=1}^N \gamma_{1,k}^H (\mathbf{h}_k^H \mathbf{x} + n_k)$$

$$= \left(\sum_{k=1}^N \gamma_{1,k}^H \mathbf{h}_k^H\right) \mathbf{x} + n$$

$$= (\mathbf{h}_1^{\text{eff}})^H \mathbf{x} + n,$$

where $\mathbf{h}_1^{\text{eff}} = \sum_{k=1}^N \gamma_{1,k} \mathbf{h}_k = \mathbf{H}_1 \boldsymbol{\gamma}_1$ and $n = \sum_{k=1}^N \gamma_{1,k}^H n_k$ is unit variance complex Gaussian noise because $|\boldsymbol{\gamma}_1| = 1$. Since any set of weights satisfying the unit norm can be
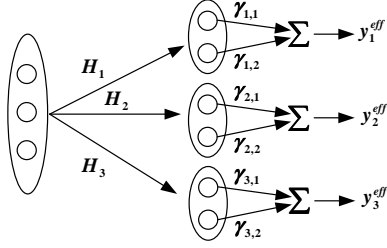
Fig. 1. Effective Channel for $M = K = 3$, $N = 2$ System

Fig. 2. Quantization procedure for a two antenna mobile

chosen, $\mathbf{h}_1^{\text{eff}}$ can be in *any* direction in the subspace spanned by $\mathbf{h}_1, \ldots, \mathbf{h}_N$. Thus, the quantization error is minimized by choosing $\mathbf{h}_1^{\text{eff}}$ to be in the direction that can be quantized best, or equivalently the direction which is closest to one of the quantization vectors. The combiner structure for a 3 user channel with $M = 3$ and $N = 2$ is shown in Fig. 1.

Let us now more formally describe the quantization process performed at the first mobile. As described in Section II-A, the quantization codebook consists of 2^B isotropically chosen unit norm vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{2^B}$. In the single receive antenna ($N = 1$, $\mathbf{H}_i = \mathbf{h}_i$) scenario, quantization is performed by choosing the quantization vector that has the smallest angle between itself and the channel vector $\mathbf{h}_1$. When $N > 1$, we compute the angle between each quantization vector and the *subspace* spanned by the N channel vectors, and pick the quantization vector that forms the smallest such angle. Alternatively, each quantization vector is projected onto the span of the N channel vectors, and the angle between the quantization vector and its projection is computed. If $\mathbf{q}_1, \ldots, \mathbf{q}_N$ forms an orthonormal basis for $\text{span}(\mathbf{h}_1, \ldots, \mathbf{h}_N)$ (easily computable using Gram-Schmidt), then the quantization is performed according to:

$$\hat{\mathbf{h}}_1 = \arg \min_{\mathbf{w} = \mathbf{w}_1, \ldots, \mathbf{w}_{2^B}} |\angle(\mathbf{w}, \text{span}(\mathbf{h}_1, \ldots, \mathbf{h}_N))| \quad (6)$$

$$= \arg \max_{\mathbf{w} = \mathbf{w}_1, \ldots, \mathbf{w}_{2^B}} \sum_{k=1}^{N} |\mathbf{w}^H \mathbf{q}_k|^2. \quad (7)$$

Let us denote the normalized projection of $\hat{\mathbf{h}}_1$ onto $\text{span}(\mathbf{h}_1, \ldots, \mathbf{h}_N)$ by the vector $\mathbf{s}_1^{\text{proj}}$. Notice that the direction specified by $\mathbf{s}_1^{\text{proj}}$ has the minimum quantization error amongst all directions in $\text{span}(\mathbf{h}_1, \ldots, \mathbf{h}_N)$. Thus, the effective channel should be chosen in this direction, i.e., we wish to choose a unit norm vector $\boldsymbol{\gamma}_1$ such that $\mathbf{h}^{\text{eff}} = \sum_{j=1}^{N} \gamma_{1,j} \mathbf{h}_j = \mathbf{H}_1 \boldsymbol{\gamma}_1$ is in the direction of the projected quantization vector $\mathbf{s}_1^{\text{proj}}$. First we find the vector $\mathbf{u} \in \mathcal{C}^{\mathbf{N}}$ such that $\mathbf{H}_1 \mathbf{u} = \mathbf{s}_1^{\text{proj}}$, and then scale to get $\boldsymbol{\gamma}_1$. Since $\mathbf{s}_1^{\text{proj}}$ is in $\text{span}(\mathbf{H}_1)$, $\mathbf{u}$ can be found by the pseudo-inverse of $\mathbf{H}_1$:

$$\mathbf{u} = \left(\mathbf{H}_1^H \mathbf{H}_1\right)^{-1} \mathbf{H}_1^H \mathbf{s}_1^{\text{proj}}, \quad (8)$$

and the coefficient vector $\boldsymbol{\gamma}_1$ is the normalized version of $\mathbf{u}$: $\boldsymbol{\gamma} = \frac{\mathbf{u}}{||\mathbf{u}||}$. It is easy to check that $||\mathbf{h}_1^{\text{eff}}|| = 1/||\mathbf{u}||$.

The quantization procedure is illustrated for a $N = 2$ channel in Fig. 2. In the figure the span of the two channel vectors is shown, along with the projection of the best quantization
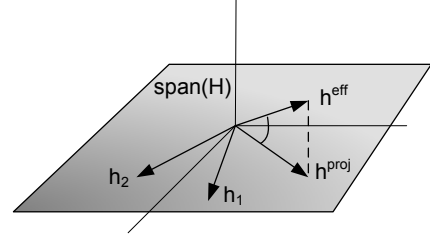
vector onto this subspace along with the subsequent angular error.

We now summarize the procedure for computing the quantization vector and the weighting vector of the i-th mobile:

1) Compute the channel quantization:

$$\hat{\mathbf{h}}_i = \arg \min_{\mathbf{w} = \mathbf{w}_1, \ldots, \mathbf{w}_{2^B}} |\angle(\mathbf{w}, \text{span}(\mathbf{H}_i))|$$

$$= \arg \max_{\mathbf{w} = \mathbf{w}_1, \ldots, \mathbf{w}_{2^B}} \sum_{k=1}^{N} |\mathbf{w}^H \mathbf{q}_k|^2. \quad (9)$$

where $\mathbf{q}_1, \ldots, \mathbf{q}_N$ is an orthonormal basis for the span of the columns of $\mathbf{H}_i$.

2) Project the quantization vector onto the span of the channel vectors:

$$\mathbf{s}_i^{\text{proj}} = \frac{\sum_{k=1}^{N} \mathbf{q}_k (\hat{\mathbf{h}}_i^H \mathbf{q}_k)}{|| \sum_{k=1}^{N} \mathbf{q}_k (\hat{\mathbf{h}}_i^H \mathbf{q}_k)||}.$$

3) Compute the weighting vector $\boldsymbol{\gamma}_i$:

$$\boldsymbol{\gamma}_i = \frac{\left(\mathbf{H}_i^H \mathbf{H}_i\right)^{-1} \mathbf{H}_i^H \mathbf{s}_1^{\text{proj}}}{|| \left(\mathbf{H}_i^H \mathbf{H}_i\right)^{-1} \mathbf{H}_i^H \mathbf{s}_1^{\text{proj}}||}. \quad (10)$$

Each mobile performs these steps, feeds back the index of its quantized channel, and then linearly combines its N received signals using weighting vector $\boldsymbol{\gamma}_i$ to get $y_i^{\text{eff}} = (\mathbf{h}_i^{\text{eff}})^H \mathbf{x} + n$ with $\mathbf{h}_i^{\text{eff}} = \mathbf{H}_i \boldsymbol{\gamma}_i$.

The proposed method finds the effective channel with the minimum quantization error without any regard to the resulting channel magnitude (i.e., $||\mathbf{h}_i^{\text{eff}}||$). This is reasonable because quantization error is the dominating factor in limited feedback downlink systems, as we later see in the sum rate analysis. However, it may be useful to later study alternatives that balance minimization of quantization error with maximization of channel magnitude.

IV. Sum Rate Analysis

The effective channel quantization procedure converts the multiple transmit, multiple receive antenna downlink channel into a multiple transmit, single receive antenna downlink channel with channel vectors $\mathbf{h}_1^{\text{eff}}, \ldots, \mathbf{h}_M^{\text{eff}}$ and channel quantizations $\hat{\mathbf{h}}_i \cdots \hat{\mathbf{h}}_M$. In fact, the transmitter need not even be aware of the number of receive antennas, since the multiple receive antennas are used only during quantization.

After receiving the quantization indices from each of the mobiles, the transmitter performs zero-forcing beamforming

(as described in Section II-B) based on the channel quantizations. The resulting SINR at the i-th receiver is given by:

$$SINR_i = \frac{\frac{P}{M}|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_i|^2}{1 + \sum_{j \neq i} \frac{P}{M}|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_j|^2}. \tag{11}$$

We are interested in the long-term average sum rate achieved in this channel, and thus the expectation of $\sum_{i=1}^M \log(1 + SINR_i)$. Since the beamforming vectors are chosen according to the ZFBF criterion based on the quantized channels, they satisfy $||\mathbf{v}_i|| = 1$ for all i and $\hat{\mathbf{h}}_i \mathbf{v}_j = 0$ for all $j \neq i$. Quantization error, however, leads to mismatch between the effective channels and their quantizations, and thus strictly positive interference terms (of the form $|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_j|^2$) in the denominator of the SINR expression.

A. Preliminary Calculations

In order to analyze the expected rate of such a system, the distribution (over the random channels and quantization codebooks) of the quantization error between $\mathbf{h}_i^{\text{eff}}$ and $\hat{\mathbf{h}}_i$ and of the effective channel must be characterized.

Lemma 1: The quantization error $\sin^2(\angle(\hat{\mathbf{h}}_i, \mathbf{h}_i^{\text{eff}}))$, is the minimum of 2^B independent beta $(M - N, N)$ random variables.

Proof: If $\mathbf{q}_1, \ldots, \mathbf{q}_N$ denote an orthonormal basis for $\text{span}(\mathbf{H}_i)$, $\cos^2(\angle(\mathbf{w}_j, \text{span}(\mathbf{H}_i))) = \sum_{k=1}^N |\mathbf{w}_j^H \mathbf{q}_k|^2$ for any quantization vector. Since the basis vectors and quantization vectors are isotropically chosen, this quantity is the squared norm of the projection of a random vector in $\mathcal{C}^M$ onto a random N-dimensional subspace, which is described by the beta distribution with parameters N and $M - N$ [9]. By the properties of the beta distribution, $\sin^2(\angle(\mathbf{w}_j, \text{span}(\mathbf{H}_i))) = 1 - \cos^2(\angle(\mathbf{w}_j, \text{span}(\mathbf{H}_i)))$ is beta $(M - N, N)$. Finally, the independence of the quantization vectors and the channels implies independence of the 2^B random variables. ∎

The following lemma and conjecture characterize the distribution of the effective channel vectors.

Lemma 2: The normalized effective channels $\frac{\mathbf{h}_1^{\text{eff}}}{||\mathbf{h}_1^{\text{eff}}||}, \ldots, \frac{\mathbf{h}_M^{\text{eff}}}{||\mathbf{h}_M^{\text{eff}}||}$ are iid isotropic vectors in $\mathcal{C}^M$.

Proof: From the earlier description of effective channel quantization, note that $\frac{\mathbf{h}_i^{\text{eff}}}{||\mathbf{h}_i^{\text{eff}}||} = \mathbf{s}_i^{\text{proj}}$, which is the projection of the best quantization vector onto $\text{span}(\mathbf{H}_i)$. Since each quantization vector is chosen isotropically, its projection is isotropically distributed within the subspace. Furthermore, the best quantization vector is chosen based solely on the angle between the quantization vector and its projection. Thus $\mathbf{s}_i^{\text{proj}}$ is isotropically distributed in $\text{span}(\mathbf{H}_i)$. Since this subspace is also isotropically distributed, the vector $\mathbf{s}_i^{\text{proj}}$ is isotropically distributed in $\mathcal{C}^M$. Independence holds due to the independence of the quantization vectors and channel realizations. ∎

Conjecture 1: The squared norm of the effective channel $\mathbf{h}_i^{\text{eff}}$ is chi-squared with $2(M - N + 1)$ degrees of freedom.

While this conjecture can be proven for the case when $N = M$ using the fact that the diagonal entries of $(\mathbf{H}_i \mathbf{H}_i^H)^{-1}$ are each inverted chi-square with two degrees of freedom when $\mathbf{H}_i$ is square [10, Theorem 3.2.12], this proof does not

yet extend to the scenario where $1 < N < M$. However, numerical results very strongly indicate that the conjecture is true for all values of N and M. The claim is trivially true when $N = 1$ because $\mathbf{h}_i^{\text{eff}} = \mathbf{h}_i$ when mobiles have a single antenna). Furthermore, it is known that $1/(\mathbf{v}^H \mathbf{H}_i^H \mathbf{H}_i)^{-1}\mathbf{v})$ is chi-square distributed with $2(M - N + 1)$ degrees of freedom for any unit norm $\mathbf{v}$ [10].

If $N = M$, there is zero quantization error but the resulting effective channels have only two degrees of freedom. This scenario is not relevant, however, because higher rates can be achieved by simply transmitting to a single user using point-to-point MIMO techniques, since such a system has the same number of spatial degrees of freedom as the downlink channel.

B. Sum Rate Performance Relative to Perfect CSIT

In order to study the effect of finite rate feedback, we compare the sum rate achieved using finite rate feedback and effective channel quantization (for $N \geq 1$), denoted $R_{FB}(P)$, to the sum rate achieved with perfect CSIT in an M transmit, *single* receive antenna downlink channel, denoted $R_{ZF}(P)$. We use the single receive antenna downlink with perfect CSIT as the benchmark instead of the N receive antenna perfect CSIT downlink channel because the proposed method effectively utilizes a single receive antenna per mobile for reception, and thus cannot outperform a single receive antenna downlink channel with perfect CSIT, even in the limit of an infinite number of feedback bits. Furthermore, this analysis allows us to compare the required feedback load with $N > 1$ and the proposed method to the required feedback load for downlink channels with single receive antennas, studied in [2][4].

Let us first analyze the rates achieved in a single receive antenna downlink channel using ZFBF under the assumption of perfect CSIT. If the transmitter has perfect CSIT, the beamforming vectors (denoted $\mathbf{v}_{ZF,i}$) can be chosen perfectly orthogonal to all other channels, thereby eliminating all multi-user interference. Thus, the SNR of each user is as given in (5) with zero interference terms in the denominator. The resulting average rate is given by:

$$E_{\mathbf{H}}[R_{ZF}(P)] = E_{\mathbf{H}}\left[\log\left(1 + \frac{P}{M}|\mathbf{h}_i^H \mathbf{v}_{ZF,i}|^2\right)\right].$$

Since the beamforming vector $\mathbf{v}_{ZF,i}$ is chosen orthogonal to the $(M-1)$ other channel vectors $\{\mathbf{h}_j\}_{j \neq i}$, each of which is an iid isotropic vector, the beamforming vector is also an isotropic vector, *independent* of the channel vector $\mathbf{h}_i$. Because the effective channel vectors are isotropically distributed (Lemma 2), the same is true of the beamforming vectors and effective channel vectors when the proposed method is used.

If the number of feedback bits is fixed, the rates achieved with finite rate feedback are bounded even as the SNR is taken to infinity. Thus, the number of feedback bits must be appropriately scaled in order to avoid this limitation. Furthermore, it is useful to consider the scaling of bits required to maintain a desired rate (or power) gap between perfect CSIT and limited feedback. Thus, we study the rate gap

at asymptotically high SNR, denoted as ΔR. Some simple algebra yields the following upper bound to ΔR:

$$\Delta R \triangleq \lim_{P\to\infty} E_{\mathbf{H},W}[R_{ZF}(P) - R_{LF}(P)]$$

$$\leq \lim_{P\to\infty} E_{\mathbf{H}}\left[\log\left(1 + \frac{P}{M}|\mathbf{h}_i^H \mathbf{v}_{ZF}|^2\right)\right] -$$

$$E_{\mathbf{H},W}\left[\log\left(1 + \frac{P}{M}|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_i|^2\right)\right] +$$

$$E_{\mathbf{H},W}\left[\log\left(1 + \sum_{j\neq i}\frac{P}{M}|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_j|^2\right)\right]$$

The difference of the first two terms is the rate loss due to the reduced effective channel norm (Conjecture 1) and can be computed in closed form using the expectation of the log of chi-square random variables, giving a loss of $\Delta_a \triangleq \log_2 e \sum_{l=M-N+1}^{M-1} \frac{1}{l}$. The final term is the rate loss due to the quantization error and can be upper bounded using Jensen's inequality and some of the techniques from [2][4] to give:

$$\Delta R \leq \Delta_a + \log E_{\mathbf{H},W}\left[1 + \sum_{j\neq i}\frac{P}{M}|(\mathbf{h}_i^{\text{eff}})^H \mathbf{v}_j|^2\right]$$

$$= \Delta_a + \log\left[1 + P\left(\frac{M-N+1}{M}\right)E[\sin^2(\angle(\hat{\mathbf{h}}_i, \mathbf{h}_i^{\text{eff}}))]\right]$$

We now utilize Lemma 1 to estimate the quantization error. If we let X be a beta$(M-N, N)$ random variable, the CCDF of X can be accurately approximated for $x \approx 1$ as $\Pr(X \geq x) \approx \binom{M-1}{N-1}(1-x)^{(M-N)}$. Since the quantization error is one minus the maximum of 2^B such random variables, we use extreme value theory and find x such that $\Pr(X \geq x) = 2^{-B}$ to get the following approximation for the quantization error:

$$E[\sin^2(\angle(\hat{\mathbf{h}}_i, \mathbf{h}_i^{\text{eff}}))] \approx 2^{-\frac{B}{M-N}}\binom{M-1}{N-1}^{-\frac{1}{M-N}}$$

Thus we have

$$\Delta R \approx \log_2 e \sum_{l=M-N+1}^{M-1} \frac{1}{l} +$$

$$\log\left(1 + P \cdot \left(\frac{M-N+1}{M}\right)2^{-\frac{B}{M-N}}\binom{M-1}{N-1}^{-\frac{1}{M-N}}\right)$$

If we set this quantity equal to a desired rate gap $r > 0$ and solve for the required scaling of B as a function of the SNR (in dB) we get:

$$B = \frac{M-N}{3}P_{dB} - (M-N)\log_2 c \qquad (12)$$

$$-(M-N)\log_2\left(\frac{M}{M-N+1}\right) - \log_2\binom{M-1}{N-1},$$

where $c = 2^r \cdot e^{-(\sum_{l=M-N+1}^{M-1} \frac{1}{l})} - 1$. Note that a per user rate gap of $r = 1$ bps/Hz is equivalent to a 3 dB gap in the sum rate curves. If we compute the difference between this expression and the feedback load required when $N = 1$ (given in (1)) and a 3 dB gap is desired ($r = 1$), we can get the following
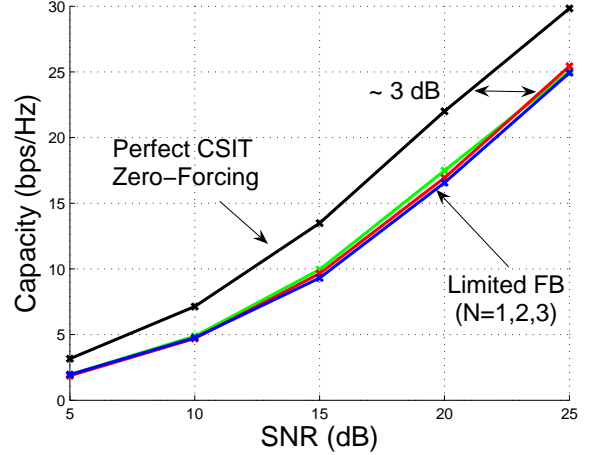

Fig. 3. Sum rate of $M = K = 6$ downlink channel

approximation for the feedback reduction as a function of the number of mobile antennas N:

$$\Delta_{FB}(N) \approx \frac{N-1}{3}P_{dB} + \log_2\binom{M-1}{N-1} - (N-1)\log_2 e.$$

For $N = 2$, the feedback savings is given by:

$$\Delta_{FB}(2) \approx \frac{1}{3}P_{dB} + \log_2(M-1) - \log_2 e.$$

The sum rate of a 6 transmit antenna downlink channel is plotted in Fig. 3. The perfect CSIT zero-forcing curve is plotted along with the rates achieved using finite rate feedback with the feedback load scaled as specified in (12) for $N = 1, 2$ and 3. Notice that the rates achieved for different numbers of transmit antennas are nearly indistinguishable, and all three curves are approximately 3 dB shifts of the perfect CSIT curve. In this system, the feedback savings at 20 dB is 7 and 12 bits, respectively, for 2 and 3 receive antennas.

REFERENCES

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[2] N. Jindal, "MIMO broadcast channels with finite rate feedback," in *Proceedings of IEEE Globecom, 2005*.

[3] P. Ding, D. Love, and M. Zoltowski, "Multiple antenna broadcast channels with partial and limited feedback," 2005, submitted to IEEE Trans. Sig. Proc.

[4] N. Jindal, "MIMO broadcast channels with finite rate feedback," March 2006, submitted to *IEEE Trans. Inform. Theory*.

[5] W. Santipach and M. Honig, "Asymptotic capacity of beamforming with limited feedback," in *Proceedings of Int. Symp. Inform. Theory*, July 2004, p. 290.

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

[7] D. Love, R. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[8] 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, no. 10, pp. 2562–2579, Oct. 2003.

[9] A. K. Gupta and S. Nadarajah, *Handbook of Beta Distribution and Its Applications*. CRC, 2004.

[10] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, 1982.